A

From Sylvester-Gallai Configurations to Rank Bounds: Improved Blackbox Identity Test for Depth-3 Circuits

Nitin Saxena, Hausdorff Center for Mathematics, Bonn
C. Seshadhri, Sandia National Labs, Livermore

We study the problem of identity testing for depth-3 circuits of top fanin k and degree d. We give a new structure theorem for such identities that improves the known deterministic $d^{k^{O(k)}}$-time blackbox identity test over rationals (Kayal & Saraf, FOCS 2009) to one that takes $d^{O(k^2)}$-time. Our structure theorem essentially says that the number of independent variables in a real depth-3 identity is very small. This theorem affirmatively settles the strong rank conjecture posed by Dvir & Shpilka (STOC 2005).

We devise various algebraic tools to study depth-3 identities, and use these tools to show that any depth-3 identity contains a much smaller *nucleus identity* that contains most of the "complexity" of the main identity. The special properties of this nucleus allow us to get near optimal rank bounds for depth-3 identities. The most important aspect of this work is relating a field-dependent quantity, the *Sylvester-Gallai rank bound*, to the rank of depth-3 identities. We also prove a high dimensional Sylvester-Gallai theorem for all fields, and get a general depth-3 identity rank bound (slightly improving previous bounds).

Categories and Subject Descriptors: F.1.1 [**Computation by abstract devices**]: Circuits; I.1.2 [**Symbolic and algebraic manipulation**]: Algebraic Algorithms; G.2.1 [**Discrete mathematics**]: Combinatorics

General Terms: Algorithms, Theory

Additional Key Words and Phrases: Chinese remaindering, combinatorial design, depth-3 circuit, ideal theory, identities, incidence geometry, Sylvester-Gallai

ACM Reference Format:
Saxena, N. and Seshadhri, C. 2012. From Sylvester-Gallai Configurations to Rank Bounds: Improved Blackbox Identity Test for Depth-3 Circuits. J. ACM V, N, Article A (January YYYY), 35 pages.
DOI = 10.1145/0000000.0000000 http://doi.acm.org/10.1145/0000000.0000000

1. INTRODUCTION

Polynomial identity testing (PIT) ranks as one of the most important open problems in the intersection of algebra and computer science. We are provided an arithmetic circuit that computes a polynomial $p(x_1, x_2, \cdots, x_n)$ over a field $\mathbb{F}$, and we wish to test if p is identically zero (in other words, if p is the zero polynomial). In the blackbox setting, we do not have access to the circuit. We are only allowed to evaluate the polynomial p at various domain points. The main goal is to devise a *deterministic* (preferably blackbox) polynomial time algorithm for PIT. [Heintz and Schnorr 1980; Kabanets and Impagliazzo 2004] and Agrawal [2005; 2006] have shown connections

A preliminary version of this paper appeared as [Saxena and Seshadhri 2010].
Author's details: Nitin Saxena, Hausdorff Center for Mathematics, Bonn, Germany ns@hcm.uni-bonn.de; C. Seshadhri (This work was done when the author was a postdoc at IBM Almaden) Sandia National Labs, Livermore, USA. The author is currently with Sandia National Laboratories, a multi-program laboratory managed and operated by Sandia Corporation, a wholly owned subsidiary of Lockheed Martin Corporation, for the U.S. Department of Energy's National Nuclear Security Administration under contract DE-AC04-94AL85000.

DOI 10.1145/0000000.0000000 http://doi.acm.org/10.1145/0000000.0000000

between deterministic algorithms for identity testing and circuit lower bounds, emphasizing the importance of this problem. For a detailed exposition, see surveys [Shpilka and Yehudayoff 2010; Saxena 2009; Agrawal and Saptharishi 2009]. Also, [Mulmuley 2011; Mulmuley 2012] discusses blackbox identity testing and the arithmetic P vs NP question, in the language of *geometric complexity theory*.

Even for the special case of depth-3 circuits (cf. Section 1.1), this question is still open. This may seem quite depressing. It is. Nonetheless, there exist concrete results that justify both our ignorance and the acceptance of results on depth-3 PIT in major publishing venues. [Agrawal and Vinay 2008] showed that an efficient blackbox identity test for depth-4 essentially leads to exponential lower bounds. The importance of depth-3 circuits is underscored by a result of [Raz 2010], who proved that strong lower bounds for depth-3 circuits imply super-polynomial lower bounds for general arithmetic formulas.

A depth-3 circuit C over a field $\mathbb{F}$ is of the form $C(x_1, \cdots, x_n) = \sum_{i=1}^{k} T_i$, where T_i (*a multiplication term*) is a product of at most d linear polynomials with coefficients in $\mathbb{F}$. We are especially interested in the case $\mathbb{F} = \mathbb{Q}$. In this section, we will just assume this unless explicitly mentioned otherwise. The size of the circuit C can be expressed in three parameters: the number of variables n, the degree d, and the *top fanin* (or the number of terms) k. Such a circuit is referred to as a $\Sigma\Pi\Sigma(k, d, n)$ circuit. PIT algorithms for depth-3 circuits were first studied by [Dvir and Shpilka 2006]. There have been many recent results in this area by [Kayal and Saxena 2007] (in the non-blackbox setting) and [Karnin and Shpilka 2008; Saxena and Seshadhri 2011a; Kayal and Saraf 2009]. Our main result is a better blackbox tester for $\Sigma\Pi\Sigma$ circuits over $\mathbb{Q}$. We get a running time of nd^{k^2}, an exponential improvement (in k) over the previous best of nd^{k^k} [Kayal and Saraf 2009]. Table I details the time complexities of previous algorithms[1].

THEOREM 1.1. *Consider circuits over $\mathbb{Q}$. There exists a deterministic blackbox algorithm for PIT on $\Sigma\Pi\Sigma(k, d, n)$ circuits, whose time complexity is poly(nd^{k^2}).*

Table I: Depth-3 blackbox PIT algorithms over $\mathbb{Q}$

Paper	Time complexity
[Karnin and Shpilka 2008]	$nd^{(2^{k^2}\log^{k-2} d)}$
[Saxena and Seshadhri 2011a]	$nd^{k^3 \log d}$
[Kayal and Saraf 2009]	$nd^{(k^k)}$
This paper	nd^{k^2}

This is the first result that gives a time complexity both polynomial in d and singly-exponential in k for $\mathbb{Q}$. This is not too far from the best *non-blackbox* algorithm for $\Sigma\Pi\Sigma$ circuits, which runs in poly(nd^k) time [Kayal and Saxena 2007]. This result closes the gap (almost) between blackbox and non-blackbox algorithms. Recently, this gap was closed (for all fields) by a different blackbox algorithm [Saxena and Seshadhri 2011b] borrowing the algebraic techniques developed in this paper.

All these results go via *rank bounds for depth-3 identities*, introduced by [Dvir and Shpilka 2006]. A depth-3 identity is simply a depth-3 circuit that computes

[1]These time complexities are actually bounds on the total number of bit operations. Also, the running times are technically polynomial in the stated times.

the zero polynomial. This is a very interesting quantity associated with these circuits, and roughly speaking, bounds the maximum number of "free variables" that can be present in a depth-3 identity. If a $\Sigma\Pi\Sigma(k,d,n)$ circuit has rank r, then there exists a linear transformation that converts this to an equivalent $\Sigma\Pi\Sigma(k,d,r)$ circuit. (This linear transformation is very easy to determine.) The remarkable insight of [Dvir and Shpilka 2006] was that the rank of *every* $\Sigma\Pi\Sigma(k,d,n)$ identity is very low. Any $\Sigma\Pi\Sigma(k,d,r)$-circuit can be completely expanded out in poly(kd^r) time. Hence, low rank bounds for identities imply efficient non-blackbox PIT algorithms.

[Karnin and Shpilka 2008] showed how small rank bounds for identities imply efficient *blackbox* PIT algorithms. This opened the door for blackbox algorithms for depth-3 PIT. Indeed, a majority of the known algorithms for this problem come as a consequence of their result. Rank bounds have also found applications in learning $\Sigma\Pi\Sigma$ circuits [Shpilka 2009; Karnin and Shpilka 2009]. Hence, the rank and file of researchers studying this problem are interested in proving small rank bounds. As mentioned earlier, we focus on the field $\mathbb{Q}$. [Dvir and Shpilka 2006] initiated this line of work by showing that the rank of a simple, minimal[2] $\Sigma\Pi\Sigma(k,d)$ identity is $2^{O(k^2)}(\log d)^{k-2}$. There are basic constructions of rank $\Omega(k)$ identities over $\mathbb{Q}$ [Dvir and Shpilka 2006]. They conjectured that the rank should be bounded by poly(k). This rank bound was improved to $O(k^3 \log d)$ by [Saxena and Seshadhri 2011a]. [Kayal and Saraf 2009] achieved a breakthrough by proving a rank bound independent of d. Their bound was $k^{O(k)}$. We finally settle the Dvir-Shpilka conjecture and show a rank bound of $O(k^2)$.

The advances of Kayal & Saraf were obtained through the use of *incidence geometry theorems*, like the famous Sylvester-Gallai theorem. This theorem states that for any set S of points in the Euclidean plane, not all collinear, there exists a line passing through exactly two points in S. Higher dimensional generalizations are called *Sylvester-Gallai theorems* (see survey [Borwein and Moser 1990]). These theorems have connections to rank bounds for depth-3 circuits. The result of [Kayal and Saraf 2009] gave an intricate combinatorial construction that converts depth-3 identities to sets of colored points in Euclidean space. This allowed the use of Sylvester-Gallai theorems to bound the rank.

Our contribution comes through a new algebraic framework for studying depth-3 identities. It allows for a much more "efficient" use of Sylvester-Gallai theorems to bound the rank. This leads to nearly optimal rank bounds. The connection between Sylvester-Gallai theorems and rank bounds is far more transparent, at the loss of some color from the theorems. Theorem 1.4 gives a simple formula that relates the depth-3 rank to Sylvester-Gallai bounds. A nice byproduct of this connection is the improvement of rank bounds over arbitrary fields. Most importantly, we develop a Chinese Remainder Theorem for depth-3 identities, inspired by techniques in [Kayal and Saxena 2007]. As we mentioned earlier, there have been some recent improvements for blackbox PIT algorithms for depth-3 circuits [Saxena and Seshadhri 2011b]. This algorithm does not go via the usual route of rank bounds, so this result does not subsume the rank bounds given in this paper. Nonetheless, one of the main ingredients for [Saxena and Seshadhri 2011b] is the Chinese Remainder Theorem given in this paper.

1.1. Definitions and results

We recall that a depth-3 circuit C over a field $\mathbb{F}$ is: $C(x_1,\ldots,x_n) = \sum_{i=1}^{k} T_i$, where T_i is a product of d_i linear polynomials $\ell_{i,j}$ over $\mathbb{F}$. For the purposes of studying identities we can assume, by homogenization, that $\ell_{i,j}$'s are linear *forms* (i.e. linear polynomials with a zero constant coefficient) and $\forall i, d_i = d$. For example, $(x_1+1)x_2 + x_3x_4$ can be

[2]These are small technical conditions that need to be imposed. We give details shortly.

homogenized via the map $x_i \mapsto x_i/z$ to get $(x_1 + z)x_2 + x_3x_4$. Obviously, it preserves non-zeroness and increases the number of useful variables by at most one. Refer to Lemma 3.5 of [Dvir and Shpilka 2006] for a formal explanation.

It will be convenient to state our results in terms of arbitrary fields.

Definition 1.2. [Dvir and Shpilka 2006]

— **Simple Circuit:** C is a *simple* circuit if there is no nonzero linear form dividing all the T_i's.
— **Minimal Circuit:** C is a *minimal* circuit if for every proper subset $S \subset [k]$, $\sum_{i \in S} T_i$ is nonzero.
— **Rank of a circuit:** The coefficients of $\ell_{i,j}$ form an n-dimensional vector over $\mathbb{F}$. The *rank* of the circuit, $\text{rk}(C)$, is defined as the rank of the set of all linear forms $\ell_{i,j}$ viewed as vectors.

The rank of a circuit can be interpreted as the minimum number of independent variables required to express C. The definition of simple and minimal circuits are used to remove certain pathological cases. The rank question is: For a simple and minimal $\Sigma\Pi\Sigma(k, d, n)$ identity over field $\mathbb{F}$, what is the maximal possible rank? A trivial upper bound on the rank (for any $\Sigma\Pi\Sigma$-circuit) is $\min(kd, n)$, since that is the total number of linear forms involved in C. A substantially smaller rank bound than kd shows that identities do not have as many "degrees of freedom" as general circuits.

Before we state our results, it will be helpful to explain *Sylvester-Gallai configurations*. A set of points S with the property that every line through two points of S passes through a third point in S is called a *Sylvester-Gallai configuration*. The famous Sylvester-Gallai theorem states that the only Sylvester-Gallai configuration in $\mathbb{R}^2$ is a set of collinear points. This basic theorem about point-line incidences was extended to higher dimensions [Hansen 1965; Bonnice and Edelstein 1967]. We define the notion of *Sylvester-Gallai rank bounds*. This is a clean and convenient way of expressing these theorems.

Definition 1.3. Let S be a finite subset of the *projective space* $\mathbb{FP}^n$. Alternately, S is a set of non-zero vectors in $\mathbb{F}^{n+1}$ without *multiples*: No two vectors in S are scalar multiples of each other[3]. Suppose, for every set $V \subset S$ of k linearly independent vectors, the linear span of V contains at least $k + 1$ vectors of S. Then, the set S is said to be *SG_k-closed*.

The largest possible rank of an SG_k-closed set of at most m vectors in $\mathbb{F}^n$ (for any n) is denoted by $\text{SG}_k(\mathbb{F}, m)$.

The Sylvester-Gallai theorem states[4] that for all m, $\text{SG}_2(\mathbb{R}, m) \leq 2$. Higher dimensional analogues [Hansen 1965; Bonnice and Edelstein 1967] can be interpreted to say $\text{SG}_k(\mathbb{R}, m) \leq 2(k-1)$. Our main theorem is a simple, clean expression of how Sylvester-Gallai influences identities. We state this for general fields.

THEOREM 1.4 (FROM SG_k TO RANK). *Let $|\mathbb{F}| > d$. The rank of a simple and minimal $\Sigma\Pi\Sigma(k, d)$ identity over $\mathbb{F}$ is at most $2k^2 + k \cdot \text{SG}_k(\mathbb{F}, d)$.*

Remarks: (1) We make the field assumption due to the technicalities of the projective space. If $\mathbb{F}$ is small, then we choose an extension $\mathbb{F}' \supset \mathbb{F}$ of size $> d$ and get a rank bound

[3]When $|\mathbb{F}| > |S|$, such an S is, wlog, a subset of distinct vectors with first coordinate 1.
[4]To see this, take an SG_2-closed set S of vectors. Think of each vector being represented by an infinite line through the origin, hence giving a set S in the projective space. Take a 2-dimensional plane P not passing through the origin and take the set of intersection points I of the lines in S with P. Observe that the coplanar points I have the property that a line passing through two points of I passes through a third point of I.

with $\mathrm{SG}_k(\mathbb{F}', d)$. Clearly, $\mathrm{SG}_k(\mathbb{F}, d) \leq \mathrm{SG}_k(\mathbb{F}', d)$ (where the inequality can be strict in certain cases, eg. $\mathbb{R} \subset \mathbb{C}$). Thus, an upper bound on $\mathrm{SG}_k(\mathbb{F}', d)$ implies one on $\mathrm{SG}_k(\mathbb{F}, d)$, which suffices for our applications.

(2) A relevant special case is that of syntactically *multilinear* $\Sigma\Pi\Sigma(k, d)$ identities [Shpilka and Volkovich 2008; Shpilka and Volkovich 2009; Karnin et al. 2010]. Since each product gate here comprises of linear forms over disjoint variables, it rules out the existence of any nontrivial *closed set* in Theorem 2.9. Thus, we can save on the contribution arising from $\mathrm{SG}_k(\mathbb{F}, d)$ to the rank bound. Finally, we get: The rank of a simple, minimal, multilinear $\Sigma\Pi\Sigma(k, d)$ identity over $\mathbb{F}$ is at most $2k^2$.

A direct application of the $SG_k(\mathbb{R}, m)$ bound yields an almost optimal rank bound for real depth-3 identities. For completeness, we state the exact rank bound obtained. We have a slightly stronger version (Theorem 2.11) of the above theorem that gives better constants.

THEOREM 1.5 (DEPTH-3 RANK BOUNDS). *Let C be a $\Sigma\Pi\Sigma(k, d)$ circuit, over field $\mathbb{R}$, that is simple, minimal and zero. Then, rk$(C) < 3k^2$.*

As discussed before, an application of this result to Lemma 4.10 of [Karnin and Shpilka 2008] gives an efficient deterministic blackbox identity test for $\Sigma\Pi\Sigma(k, d, n)$ circuits over $\mathbb{Q}$. Formally, we get the following *hitting set generator* for $\Sigma\Pi\Sigma$ circuits with real coefficients.

COROLLARY 1.6 (BLACKBOX PIT OVER $\mathbb{Q}$). *There is a deterministic algorithm that takes as input a triple (k, d, n) of natural numbers and in time poly(nd^{k^2}), outputs a hitting set $\mathcal{H} \subset \mathbb{Z}^n$ with the following properties:*

1) Any $\Sigma\Pi\Sigma(k, d, n)$ circuit C over $\mathbb{R}$ computes the zero polynomial iff $\forall a \in \mathcal{H}$, $C(a) = 0$.
2) $\mathcal{H}$ has at most poly(nd^{k^2}) points.
3) The total bit-length of each point in $\mathcal{H}$ is poly$(kn \log d)$.

1.1.1. Other fields. What about other fields? The rank bounds of [Dvir and Shpilka 2006; Saxena and Seshadhri 2011a] hold for arbitrary fields, whereas the rank bound of [Kayal and Saraf 2009] holds only for $\mathbb{R}$. It has been observed that for finite fields, the rank of an $\Sigma\Pi\Sigma$ identity can be as large as $\Omega(k \log d)$ [Kayal and Saxena 2007; Saxena and Seshadhri 2011a]. Hence, the $O(k^3 \log d)$ bound proved by [Saxena and Seshadhri 2011a] is almost optimal. As a small bonus, we give a slight improvement upon this bound using our approach. This requires Sylvester-Gallai theorems over arbitrary fields, an interesting question in itself. It was shown that $\mathrm{SG}_2(\mathbb{C}, m) \leq 3$ [Elkies et al. 2006], and certain lower bounds for locally decodable codes implied $\mathrm{SG}_2(\mathbb{F}, m) = O(\log m)$. (Concretely, Corollary 2.9 of [Dvir and Shpilka 2006] can be used to prove that $\mathrm{SG}_2(\mathbb{F}, m) = O(\log m)$. This is an extension of theorems in [Goldreich et al. 2002] that prove this for $\mathbb{F}_2$.) Other than this, nothing was previously known. One of our auxiliary theorems gives a high-dimensional Sylvester-Gallai bound for all fields. Applying the stronger version of Theorem 1.4, we get our rank bound. [Bhattacharyya et al. 2011] subsequently proved new 2-dimensional Sylvester-Gallai bounds, for finite fields, that are optimal for small p.

THEOREM 1.7 (SG_k FOR ALL FIELDS). *For any field $\mathbb{F}$ and $k, m \in \mathbb{N}^{>1}$, $SG_k(\mathbb{F}, m) \leq (k-1)\log_2(2m)$.*

Let C be a $\Sigma\Pi\Sigma(k, d)$ circuit, over an arbitrary field $\mathbb{F}$, that is simple, minimal and zero. Then, rk$(C) < 3k^2 + \frac{k^2}{4} \lg d$.

We provide a construction showing that $\mathrm{SG}_k(\mathbb{F}_p, m) = \Omega(k \log_p(m/k))$. Recently, it has been shown that $\mathrm{SG}_2(\mathbb{F}_p, m) = O(\mathbf{poly}(p) + \log_p m)$ (Corollary 1.3 of [Bhattacharyya et al. 2011]).

1.2. History

And now, for a brief history of PIT algorithms. The first randomized polynomial time PIT algorithm, which was a blackbox algorithm, was given (independently) by [Schwartz 1980; Zippel 1979; DeMillo and Lipton 1978]. Randomized algorithms that use less randomness were given by [Chen and Kao 2000; Lewin and Vadhan 1998; Agrawal and Biswas 2003]. [Klivans and Spielman 2001] observed that even for depth-3 circuits of bounded top fanin, deterministic identity testing was open. Progress towards this was first made by the quasi-polynomial time algorithm of [Dvir and Shpilka 2006]. The problem was resolved by a polynomial time algorithm given by [Kayal and Saxena 2007], with a running time exponential in the top fanin. Both these algorithms were non-blackbox. As for blackbox algorithms, the authors are quite sure that the reader has heard enough history. Identity tests are known only for special depth-4 circuits [Arvind and Mukhopadhyay 2010; Saxena 2008], Shpilka and Volkovich [2008; 2009], [Karnin et al. 2010; Saraf and Volkovich 2011; Anderson et al. 2011; Beecken et al. 2011; Saha et al. 2012]. Recently, [Agrawal et al. 2012] presents a unified approach to study diverse circuit restrictions, by generalizing the notion of rank and employing *Jacobian* techniques. [Agrawal and Vinay 2008] showed that an efficient blackbox identity test for depth-4 circuits will actually give a quasi-polynomial time blackbox test, and exponential lower bounds, for circuits of *all depths* that compute *low degree* polynomials. Thus, understanding depth-3 identities seems to be a natural first step towards the goal of PIT.

2. PROOF OUTLINE, IDEAS, AND ORGANIZATION

Our proof of the rank bound comprises of several new ideas. Initially, we will not provide any proofs. Instead we will only provide the intuition and the overall argument. The full proof of Theorem 1.4 is extremely technical, requires many definitions, and involves many algebraic arguments. Our attempt is to first convey with main ideas without getting into too much formalism or mathematical details. We describe all the major milestones, many of which are interesting in their own right. Indeed, it is the authors' opinion that the reader has little to gain from simply reading the detailed proofs without getting the essence of the ideas.

The intuition portion is divided into three subsections, each dealing with a separate component of the final proof. Each portion proves an interesting structural theorem. The three notions that are crucially used and developed are: ideal Chinese remaindering, ideal matchings, and Sylvester-Gallai rank bounds. Related notions have appeared (in some form) in the works of [Kayal and Saxena 2007; Saxena and Seshadhri 2011a; Kayal and Saraf 2009] respectively, to prove different kinds of results. The first two steps set up the algebraic framework and prove theorems that hold for all fields. The third step is where the Sylvester-Gallai theorems are brought in.

The main result, Theorem 1.5, will be proven in Sections 4, 5, and 6. Each section will take up one of the three notions listed above. For each of these sections, there will be a subsection here that provides the intuitive ideas. Section 7 will give the proof of Theorem 1.7. We have some preliminary technical lemmas that are collected in Section 3, preceding the main sections. The proofs for these are not provided there, and instead moved to the end in Section 8. We have done this because we feel the proofs are not representative of our contributions, and moreover have little to do with depth-3

circuits specifically. Nonetheless, we will reference them quite heavily throughout the different proofs. A few "secondary" technical and algebraic lemmas will appear in the main sections, and their proofs will also be deferred to Section 8.

2.1. Notation and definitions

We will denote the set $\{1,\ldots,n\}$ by $[n]$, and $\{a,\ldots,n\}$ by $[a,n]$. We fix the base field to be $\mathbb{F}$, so the circuits compute multivariate polynomials in the *polynomial ring* $\mathcal{R} := \mathbb{F}[x_1,\ldots,x_n]$. We use $\mathbb{F}^*$ to denote $\mathbb{F}\setminus\{0\}$.

A *linear form* is a linear polynomial in $\mathcal{R}$ with zero constant term. We will denote the set of all linear forms by $L(\mathcal{R}) := \{\sum_{i=1}^{n} a_i x_i \mid a_1,\ldots,a_n \in \mathbb{F}\}$, and the nonzero ones by $L(\mathcal{R})^*$. Clearly, $L(\mathcal{R})$ is a vector (or linear) space over $\mathbb{F}$ and that will be quite useful. Much of what we do shall deal with *multi*-sets of linear forms (sometimes polynomials in $\mathcal{R}$), equivalence classes inside them, and various maps across them. A *list* of linear forms is a multi-set of forms with an arbitrary order associated with them. The actual ordering is unimportant: We will heavily use maps between lists, and the ordering allows us to define these maps unambiguously. The usual set operations between lists can be naturally defined.

Our analysis requires various ideal-theoretic notions. One can think of ideals (for our purposes) as some kind of generalization of linear subspaces.

Definition 2.1. We collect some important definitions, mostly from [Saxena and Seshadhri 2011a]:

[Ideal] An ideal is an additive subgroup of $\mathcal{R}$ closed under multiplication by elements in $\mathcal{R}$. The ideal *generated* by the set $S \subseteq \mathcal{R}$ is the set $\{\sum_{s\in S} s f_s | f_s \in \mathcal{R}\}$. The ideal generated by the elements $s_1, s_2, \ldots$ is denoted by $\langle s_1, s_2, \ldots\rangle$.

[Multiplication term, $L(\cdot)$ & $M(\cdot)$] A *multiplication term* f is an expression in $\mathcal{R}$ given as, $f := c\cdot\prod_{\ell\in S}\ell$, where $c\in\mathbb{F}^*$ and S is a list of nonzero linear forms. The *list of linear forms in* f, $L(f)$, is just the list S. For a list S of linear forms we define the *multiplication term of* S, $M(S)$, as $\prod_{\ell\in S}\ell$. (Conventionally, $L(c)=\emptyset$ and $M(\emptyset)=1$.)

[Forms in a Circuit] We will represent a $\Sigma\Pi\Sigma(k,d)$ circuit C as a sum of k multiplication terms of degree d, $C=\sum_{i=1}^{k} T_i$. The list of *linear forms occurring in* C is $L(C) := \bigcup_{i\in[k]} L(T_i)$. Note that $L(C)$ is a list of size exactly kd. The *rank of* C, $\mathrm{rk}(C)$, is just the number of linearly independent linear forms in $L(C)$. (For the purposes of this paper T_i's are given in circuit representation and thus the list $L(T_i)$ is unambiguously defined from C.)

[Similar forms] For any two polynomials $f,g\in\mathcal{R}$, f is *similar to* g if there exists $c\in\mathbb{F}^*$ such that $f=cg$. We say f *is similar to* g *mod* I, for some ideal I of $\mathcal{R}$, if there exists $c\in\mathbb{F}^*$ such that $f\equiv cg (\mathrm{mod}\ I)$. Note that "similarity mod I" is an equivalence relation and partitions any list of polynomials into equivalence classes.

[Span sp$(\cdot)$] For any $S\subseteq L(\mathcal{R})$ we let $\mathrm{sp}(S)\subseteq L(\mathcal{R})$ be the *linear span* of the linear forms in S over the field $\mathbb{F}$. (Conventionally, $\mathrm{sp}(\emptyset)=\{0\}$.)

[Matchings] Let U,V be lists of linear forms and I be a subspace of $L(\mathcal{R})$. An *I-matching π between* U,V is a bijection π between lists U,V such that: For all $\ell\in U$, $\pi(\ell)\in\mathbb{F}^*\ell + I$. (In particular, $\pi: U\to V$ satisfies the property that ℓ and $\pi(\ell)$ are similar mod I.)

When f,g are multiplication terms, an *I-matching between* f,g means an I-matching between $L(f), L(g)$.

2.2. Step 1: Chinese Remaindering for $\Sigma\Pi\Sigma$-circuits and Matchings

The first step involves a generalization of the Chinese Remainder Theorem (CRT) for depth-3 circuits (Theorem 4.6). This formalizes the white-box algorithm of

[Kayal and Saxena 2007] as a CRT over very specific ideals generated using the forms of $L(C)$. They also discuss their algorithm in terms of Chinese Remaindering. We distill all of it down to a single theorem, and give a new self-contained proof. A formal discussion of this theorem will require introducing many new definitions, so we will skip that and discuss matchings.

The CRT is used to prove that all multiplication terms of a minimal $\Sigma\Pi\Sigma$ identity can be matched by a low rank space K, spanned by "few" linear forms in $L(\mathcal{R})$.

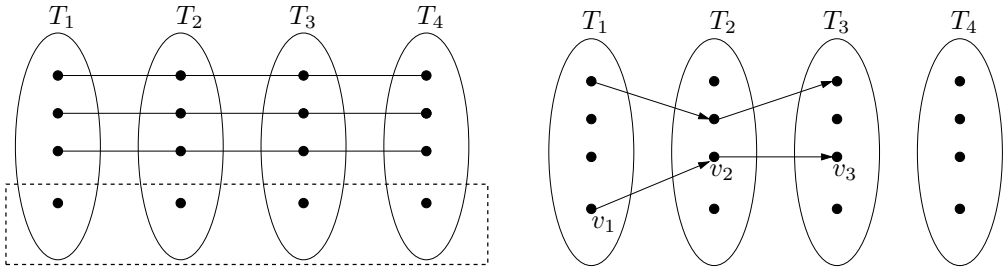
THEOREM 2.2 (MATCHING-NUCLEUS). *Let $C = T_1 + \cdots + T_k$ be a $\Sigma\Pi\Sigma(k,d)$ circuit that is minimal and zero. Then there exists a linear subspace K of $L(\mathcal{R})$ such that:*
1) $\mathbf{rk}(K) < k^2$.
2) $\forall i \in [k]$, there is a K-matching π_i between T_1, T_i.

The idea of matchings within identities was first introduced in [Saxena and Seshadhri 2011a], but nothing as powerful as this theorem was proven. This theorem gives us a space of small rank, *independent of* d, that contains most of the "complexity" of C. All forms in C outside K are just mirrored in the various terms. This starts connecting the algebra of depth-3 identities to a combinatorial structure. Indeed, the graphical picture (explained in detail below) that this theorem provides gives an intuitive picture of these identities.

We now provide a very informal description of the ideas involved in Step 1.

Definition 2.3 (*mat-nucleus*). Let C be a minimal $\Sigma\Pi\Sigma(k,d)$ identity. A linear subspace K given by Theorem 2.2 is called a *mat-nucleus of* C.

The notion of mat-nucleus is easier to see in the representation of the $\Sigma\Pi\Sigma(4,d)$ circuit $C = \sum_{i\in[4]} T_i$ given in Figure 1a. The four bubbles refer to the four multiplication terms of C and the points inside the bubbles refer to the linear forms in the terms. The proof of Theorem 2.2 gives the mat-nucleus as the space generated by the linear forms in the dotted box. The linear forms not in the mat-nucleus lie "above" the mat-nucleus and are all (mat-nucleus)-matched, i.e. $\forall \ell \in (L(T_1) \setminus \text{mat-nucleus})$, there is a form similar to ℓ modulo the mat-nucleus in each $(L(T_i) \setminus \text{mat-nucleus})$. Thus the essence of Theorem 2.2 is: The mat-nucleus part of the terms of C has low rank k^2, while the part of the terms above mat-nucleus all look "similar".



(a) Mat-nucleus (b) Paths

Fig. 1

Proof Idea for Theorem 2.2. As a full disclosure to the reader, we declare that almost all statements and definitions given in the next few paragraphs are false. Nonetheless, they convey the right idea.

As mentioned earlier, the key insight in the construction of mat-nucleus is a reinterpretation of the non-blackbox identity test of [Kayal and Saxena 2007] as a structural result for $\Sigma\Pi\Sigma$ identities. Let a *path* be a sequence of forms $(v_1, v_2, \ldots, v_{k-1})$, where

$v_i \in T_i$. The path also generates an associated *path-ideal* $\langle v_1, v_2, \ldots \rangle$. Paths are depicted in Figure 1b. Quite naturally, one can think of it graphically as a path that intersects each term (except for T_k) exactly once.

Roughly speaking, [Kayal and Saxena 2007] showed that $C = 0$ iff for every *path* (v_1, v_2, v_3) (where $v_i \in L(T_i)$): $T_4 \equiv 0 (\text{mod } v_1, v_2, v_3)$ or in ideal terms, $T_4 \in \langle v_1, v_2, v_3 \rangle$. Thus, it is enough to go through all the d^3 paths to certify the zeroness of C. This is why the time complexity of the identity test of [Kayal and Saxena 2007] is dominated by d^k. This implies a variant of the Chinese Remainder Theorem. If C is zero modulo all path ideals, then C is identically zero. More importantly for our result, if C is non-zero, there must exist a path-ideal modulo which C is non-zero. This is the essence of Theorem 4.6 (incidentally, this is the key structural property that our later work [Saxena and Seshadhri 2011b] builds on).

If we are given a $\Sigma\Pi\Sigma(4, d)$ identity C which is *minimal*, then we know that $T_1 + T_2 + T_3 \neq 0$. Thus, by applying the above interpretation of [Kayal and Saxena 2007] to $T_1 + T_2 + T_3$ we will get a path (v_1, v_2) such that $T_3 \notin \langle v_1, v_2 \rangle$. Since $C = 0$, $T_3 + T_4 \equiv 0 (\text{mod } v_1, v_2)$. But if T_4 is in $\langle v_1, v_2 \rangle$ then so will be T_3. Hence, $T_3, T_4 \not\equiv 0 (\text{mod } v_1, v_2)$. We deduce that $T_3 \equiv -T_4 (\text{mod } v_1, v_2)$ is a nontrivial congruence and this gives a $\langle v_1, v_2 \rangle$-matching between T_3, T_4 (see Lemma 3.5). By repeating this argument with a different permutation of the terms we could match different terms (by a different ideal), and finally we expect to match all the terms (by the union of the various ideals).

This argument has numerous technical problems, the most important one being that it does not really work. But all issues can be taken care of by suitable algebraic generalizations. A major stumbling block is the presence of *repeating* forms. It could happen that $(\text{mod } v_1), v_2$ occurs in many terms, or in the same term with a higher power. The most important tool developed is an ideal version of Chinese remaindering that forces us to consider not just linear forms v_1, v_2, but multiplication terms v_1, v_2 dividing T_1, T_2 respectively. We give the full proof in Section 4.

2.3. Step 2: Certificate for Linear Independence of Gates

Theorem 2.2 gives us a space K of rank $< k^2$ that matches T_1 to each term T_i. We increase the rank of this space to make it have a much stronger property. Consider $L_K(T_i) := L(T_i) \cap K$ and let K_i be the product of these forms. Formally, $K_i = M(L_K(T_i))$; it is the sub-product of forms that belong to K. Consider any index set $\mathcal{I}$ such that $\{T_i | i \in \mathcal{I}\}$ is linearly independent (i.e. $\nexists$ nonzero $\overline{\beta}$ s.t. $\sum_{i \in \mathcal{I}} \beta_i T_i = 0$). Then, we want the corresponding $\{K_i | i \in \mathcal{I}\}$ to be also linearly independent.

Because the space K matches T_1 to each T_i, the corresponding multiplication terms K_i, $i \in [k]$, themselves form a $\Sigma\Pi\Sigma(k, d')$ identity. Formally, $C' = \sum_{i \in [k]} \alpha_i K_i$ for some α_i's in $\mathbb{F}^*$ (see Lemma 5.2) is an identity. This "sub-identity" C' is called the nucleus of C. In some sense, it contains most of the complexity of C. It mirrors the linear dependencies (or lack thereof) among the T_i's, and matches all of them.

THEOREM 2.4 (NUCLEUS). *Let $C = \sum_{i \in [k]} T_i$ be a minimal $\Sigma\Pi\Sigma(k, d)$ identity and let $\{T_i | i \in \mathcal{I}\}$ be a maximal set of linearly independent terms ($1 \leq k' := |\mathcal{I}| < k$). Then there exists a linear subspace K of $L(\mathcal{R})$ such that:*

1) $\mathbf{rk}(K) < 2k^2$.
2) $\forall i \in [k]$, there is a K-matching π_i between T_1, T_i.
3) (Define $\forall i \in \mathcal{I}$, $K_i := M(L_K(T_i))$.) The terms $\{K_i | i \in \mathcal{I}\}$ are linearly independent.

Definition 2.5 (*nucleus*). Let C be a minimal $\Sigma\Pi\Sigma(k, d)$ identity. A linear subspace K given by Theorem 2.4 is called a *nucleus of* C. By Lemma 5.2, the subspace K induces an identity $C' = \sum_{i \in [k]} \alpha_i K_i$ which we call the *nucleus identity*.

Proof Idea for Theorem 2.4. The first two properties in the theorem statement are already satisfied by mat-nucleus of C. So we incrementally add linear forms to the space mat-nucleus till it satisfies property (3) and becomes the nucleus. The addition of linear forms is guided by the ideal version of Chinese remaindering. Suppose the terms T_1, T_2, T_3 are linearly independent, so $C' = T_1 + T_2 + T_3 \neq 0$. By Chinese Remaindering, there exists a path-ideal I such that $C' \neq 0 \pmod{I}$. This implies there are forms $v_1 \in L(T_1)$, $v_2 \in L(T_2)$ such that $C' \neq 0 \pmod{\langle v_1, v_2 \rangle}$. We have that $T_2 \notin \langle v_1 \rangle$ and $T_3 \notin \langle v_1, v_2 \rangle$. We now add these forms v_1, v_2 to the space mat-nucleus, and call the new space K. We can show that the new K_1, K_2, K_3 are now linearly independent.

Not surprisingly, the above argument has numerous technical problems. But it can be made to work by careful applications of the ideal version of Chinese remaindering. We give the full proof in Section 5.

2.4. Step 3: Invoking Sylvester-Gallai Theorems

As explained in Section 1.1, we rephrase the standard Sylvester-Gallai theorems in terms of *Sylvester-Gallai closure* and *rank bounds* (Definition 1.3). We state and prove the first ever general Sylvester-Gallai bound for all fields. The original version of this paper [Saxena and Seshadhri 2010] had a somewhat involved proof, but a much simpler one was suggested by [Saks 2010]. The proof with a detailed discussion is given in Section 7.

THEOREM 2.6 (GENERAL SYLVESTER-GALLAI). *For any field $\mathbb{F}$ and $k, m \in \mathbb{N}^{>1}$, $SG_k(\mathbb{F}, m) \leq (k-1)\lg 2m$.*

The following definition is very helpful in applying Sylvester-Gallai rank bounds to our scenario.

Definition 2.7 (*SG operator*). **[$\mathbf{SG}_k(\cdot)$]** Let $k, m \in \mathbb{N}^{>1}$. Suppose a set $S \subseteq \mathbb{FP}^n$ has rank greater than $\mathrm{SG}_k(\mathbb{F}, m)$ (where $|S| \leq m$). Then, by definition, S is not SG_k-closed. The *k-dimensional Sylvester-Gallai operator* $\mathrm{SG}_k(S)$ (i.e. applied on S) returns a set of k linearly independent vectors V in S whose span has no point in $S \setminus V$.

Let C be a simple and *strongly* minimal $\Sigma\Pi\Sigma(k, d)$ identity (i.e. $T_1, \ldots, T_{k-1}$ are linearly independent). Theorem 2.4 gives us a nucleus K, of rank $< 2k^2$, that matches T_1 to each term T_i. As seen in Step 2, if we look at the corresponding multiplication terms $K_i := M(L_K(T_i))$, $i \in [k]$, then they again form a $\Sigma\Pi\Sigma(k, d')$ "nucleus identity" $C' = \sum_{i \in [k]} \alpha_i K_i$, for some α_i's in $\mathbb{F}^*$, which is simple and strongly minimal. Define the *non-nucleus part* of T_i as $L^c_K(T_i) := L(T_i) \setminus K$, for all $i \in [k]$ (c in the exponent annotates "complement", since $L(T_i) = L_K(T_i) \sqcup L^c_K(T_i)$). What can we say about the rank of $L^c_K(T_i)$?

Define the non-nucleus part of C as $L^c_K(C) := \bigcup_{i \in [k]} L^c_K(T_i)$. Our goal in Step 3 is to bound $\mathrm{rk}(L^c_K(C) \bmod K)$ by $2k$ when the field is $\mathbb{R}$. This will give us a rank bound of $\mathrm{rk}(K) + \mathrm{rk}(L^c_K(C) \bmod K) < (2k^2 + 2k)$ for simple and strongly minimal $\Sigma\Pi\Sigma(k, d)$ identities over $\mathbb{R}$. The proof is mainly combinatorial, based on higher dimensional Sylvester-Gallai theorems and a property of set partitions, with a sprinkling of algebra.

We apply the SG_k operator not directly on the forms in $L(C)$ but on a suitable truncation of those forms. So we need another definition.

Definition 2.8 (*Non-K rank*). Let K be a linear subspace of $L(\mathcal{R})$. Then $L(\mathcal{R})/K$ is again a linear space (the *quotient space*). Let S be a list of forms in $L(\mathcal{R})$. The *non-K rank of S* is defined to be $\mathrm{rk}(S \bmod K)$ (i.e. the rank of S when viewed as a subset of $L(\mathcal{R})/K$).

Let C be a $\Sigma\Pi\Sigma(k,d)$ identity with nucleus K. The non-K rank of the non-nucleus part $L^c_K(T_i)$ is called the *non-nucleus rank of* T_i. Similarly, the non-K rank of the non-nucleus part $L^c_K(C)$ is called the *non-nucleus rank of* C.

We give an example to explain the non-K rank. Let $R := \mathbb{F}[z_1, \cdots, z_n, y_1, \cdots, y_m]$. Suppose $K = \mathrm{sp}(z_1, \cdots, z_n)$ and $S \subset L(R)$. We can take any element ℓ in S and simply drop all the z_i terms, i.e. 'truncate' the z-part of ℓ. This gives a set of linear forms over the y variables. The rank of these is the non-K rank of S, which we need to bound to prove our final bound.

We are now ready to state the theorem that is proved in Step 3. It basically shows a neat relationship between the non-nucleus part and Sylvester-Gallai.

THEOREM 2.9 (BOUND FOR SIMPLE, STRONGLY MINIMAL IDENTITIES). *Let $|\mathbb{F}| > d$. The non-nucleus rank of a simple and strongly minimal $\Sigma\Pi\Sigma(k,d)$ identity over $\mathbb{F}$ is at most $SG_{k-1}(\mathbb{F},d)$. More specifically, (for nucleus K) the vectors in $L(C) \setminus K$ form an SG_{k-1}-closed set*[5].

Observe that this theorem together with Theorem 2.4 gives a complete structure theorem for strongly minimal depth-3 identities. One can make suitable claims for identities that are not strongly minimal. Essentially, we just take a subset of linearly independent terms, say $T_1, \ldots, T_{k'}$, that form a basis for $\{T_i | i \in [k]\}$. We can now construct strongly minimal identities using these terms and apply the above theorem. Section 6.4 deals with this case and bounds the non-nucleus rank for all simple, minimal identities. Specifically, we get the following.

Definition 2.10 (*Independent-fanin*). Let $C = \sum_{i\in[k]} T_i$ be a $\Sigma\Pi\Sigma(k,d)$ circuit. The *independent-fanin* of C, ind-fanin(C), is defined to be the size of the maximal $\mathcal{I} \subseteq [k]$ such that $\{T_i | i \in \mathcal{I}\}$ are linearly independent polynomials[6].

We now state the following stronger version of the main theorem.

THEOREM 2.11 (FINAL BOUND). *Let $|\mathbb{F}| > d$. The rank of a simple, minimal $\Sigma\Pi\Sigma(k,d)$, independent-fanin k', identity is at most $2k^2 + (k-k') \cdot SG_{k'}(\mathbb{F},d)$.*

Remark: In particular, the rank of a simple, minimal $\Sigma\Pi\Sigma(k,d)$ identity over reals is at most $2k^2+(k-k')\cdot \mathrm{SG}_{k'}(\mathbb{R},d) \le 2k^2+(k-k')2(k'-1) < 3k^2$, proving the main theorem over reals. Likewise, for any $\mathbb{F}$, we get the rank bound of $2k^2 + (k-k') \cdot \mathrm{SG}_{k'}(\mathbb{F},d) \le 2k^2 + (k-k')(k'-1)\lg 2d \le 2k^2 + \frac{k^2}{4}\lg 2d < 3k^2 + \frac{k^2}{4}\lg d$, proving the main theorem.

Proof Idea for Theorem 2.9. We treat the non-nucleus part of the term T_1. Each form can be thought of as a point in an appropriate high-dimensional space. We essentially construct a proof by contradiction. Assuming the non-nucleus rank is more than $\mathrm{SG}_k(\mathbb{F},d)$, we apply the SG_k-operator on $L^c_K(T_1)$. The tuple obtained is used to elicit a contradiction. Modulo the nucleus, all multiplication terms look essentially the same (i.e. $\mathrm{rk}(T_1 \bmod K) = \mathrm{rk}(C \bmod K)$), so it suffices to focus attention on just one of them. Hence, we apply the SG_k-operator on a single multiplication term.

Assume C is a simple, strongly minimal $\Sigma\Pi\Sigma(k,d)$ identity with terms $\{T_i | i \in [k]\}$ and let K be its nucleus given by Step 2. It will be convenient for us to fix a linear form $y_0 \in L(\mathcal{R})^*$ and a subspace U of $L(\mathcal{R})$ such that we have the following *orthogonal* vector space decomposition $L(\mathcal{R}) = \mathbb{F}y_0 \oplus U \oplus K$. This means for any form $\ell \in L(\mathcal{R})$, there is a unique way to express $\ell = \alpha y_0 + u + v$, where $\alpha \in \mathbb{F}$, $u \in U$ and $v \in K$. Furthermore, we will assume wlog that for every form $\ell \in L^c_K(T_1)$ the corresponding α

[5]Technically, we must convert $L(C) \setminus K$ into a multiple-free set of vectors, and then make this claim.
[6]If ind-fanin$(C) = k$ then $C \neq 0$. Also, for an identity C, C is strongly minimal iff ind-fanin$(C) = k-1$.

is nonzero, i.e. each form in $L^c_K(T_1)$ is *monic* wrt y_0 (see Lemma 6.3). Technically, we do not need the extra variable y_0 and can work in a projective space. Nonetheless, it makes the presentation easier.

Definition 2.12 (*trun*$(\cdot)$). Fix a decomposition $L(\mathcal{R}) = \mathbb{F}y_0 \oplus U \oplus K$. For any form $\ell \in L^c_K(T_1)$, there is a unique way to express $\ell = \alpha y_0 + u + v$, where $\alpha \in \mathbb{F}^*$, $u \in U$ and $v \in K$.

The *truncated form trun*(ℓ) is the linear form obtained by dropping the K part and normalizing, i.e. $\text{trun}(\ell) := y_0 + \alpha^{-1}u$. (In particular, $\ell, \text{trun}(\ell)$ are similar mod K.)

Given a list of forms S we define $\text{trun}(S)$ to be the corresponding *set* (thus no repetitions) of truncated forms.

To be precise, we fix a basis $\{y_1, \ldots, y_{\text{rk}(U)}\}$ of U so that each form in $\text{trun}(L^c_K(T_1))$ has representation $y_0 + \sum_{i \geq 1} a_i y_i$ (a_i's $\in \mathbb{F}$). We view each such form as the *point* $(1, a_1, \ldots, a_{\text{rk}(U)})$ while applying Sylvester-Gallai on $\text{trun}(L^c_K(T_1))$. Assume, for the sake of contradiction, that the non-nucleus rank of T_1, $\text{rk}(\text{trun}(L^c_K(T_1))) > \text{SG}_{k-1}(\mathbb{F}, d)$. Therefore, $\text{SG}_{k-1}(\text{trun}(L^c_K(T_1)))$ gives $(k-1)$ linearly independent forms $\ell_1, \ldots, \ell_{k-1} \in (y_0 + U)$ whose span contains no other linear form of $\text{trun}(L^c_K(T_1))$.

For simplicity of exposition, let us fix $k = 4$, K spanned by z's, U spanned by y's and $\ell_i = y_0 + y_i$ ($i \in [3]$). Note that (by definition) $\text{trun}(\alpha y_0 + \sum_i \alpha_i z_i + \sum_i \beta_i y_i) = y_0 + \sum_i \frac{\beta_i}{\alpha} y_i$. We want to derive a contradiction using the SG_3-operator output $(y_0+y_1, y_0+y_2, y_0+y_3)$ and the fact that C is a simple, strongly minimal $\Sigma\Pi\Sigma(4, d)$ identity. Consider the setting given in Figure 2. (The circuit given is not identically zero, but it helps to explain our argument.) Suppose the linear forms in C that are similar to a form in $\{y_0 + y_i + K | i \in [3]\}$ are exactly those depicted in the figure. All forms within a row are K-matched. We would like to find forms $\ell'_1, \ell'_2, \ell'_3$ with the following properties: (1) $\ell'_i \equiv c_i \ell_i (\text{mod } K)$ (for some constant c_i). (2) There exists some j such that no ℓ'_i divides T_j but for each T_l ($l \neq j$), some ℓ'_i divides T_l. In this situation, we can choose $\ell'_1 = y_0 + y_1 + z_1$, $\ell'_2 = y_0 + y_2 + z_2$, and $\ell'_3 = -y_0 - y_3 + z_2$. None of these divides T_4. Observe that the triple $(y_0 + y_1 + z_1, y_0 + y_2 + z_2, y_0 + y_3 + z_1)$ does not satisfy these conditions, since no appropriate T_j can be found.

Take C modulo the ideal $I := \langle y_0 + y_1 + z_1, y_0 + y_2 + z_2, -y_0 - y_3 + z_2 \rangle$. It is easy to see that $C \equiv T_4 (\text{mod } I)$, so I "kills" the first three terms. Since C is an assumed identity, $T_4 \in I$. Thus, there is a form $\ell \in L(T_4)$ such that $\ell \in \text{sp}(\ell'_1, \ell'_2, \ell'_3)$. Since no form from ℓ'_i divides T_4, so ℓ must be a non-trivial combination of these forms. By the matching property, there exists some form $\hat{\ell} \in L(T_1)$ such that $\text{trun}(\ell) = \text{trun}(\hat{\ell})$. In other words, $\text{trun}(\ell) \in \text{trun}(L^c_K(T_1))$. But that contradicts the fact that (ℓ_1, ℓ_2, ℓ_3) forms an SG_3-tuple. This implies that the non-nucleus rank of C is at most $\text{SG}_3(\mathbb{F}, d)$.

The approach above worked because we were lucky enough to find $\ell'_1, \ell'_2, \ell'_3$ with the right properties. Can we always do this? No, because of repeating forms. Suppose, after going modulo form ℓ, the circuit looks like $x^3y + 2x^2y^2 + xy^3$. This is not simple, but *it does not have to be*. We are only guaranteed that the original circuit is simple. Once we go modulo ℓ, that property is lost. Now, the choice of *any* form kills all terms. We will use our Chinese remaindering tools and the nucleus properties to deal with this. The minimality of the nucleus identity plays a crucial role here and helps us deal with such situations. We have to prove a special theorem about partitions of $[k]$ and use strong minimality (which we did not use in the above sketch). The full proof is given in Section 6.1.

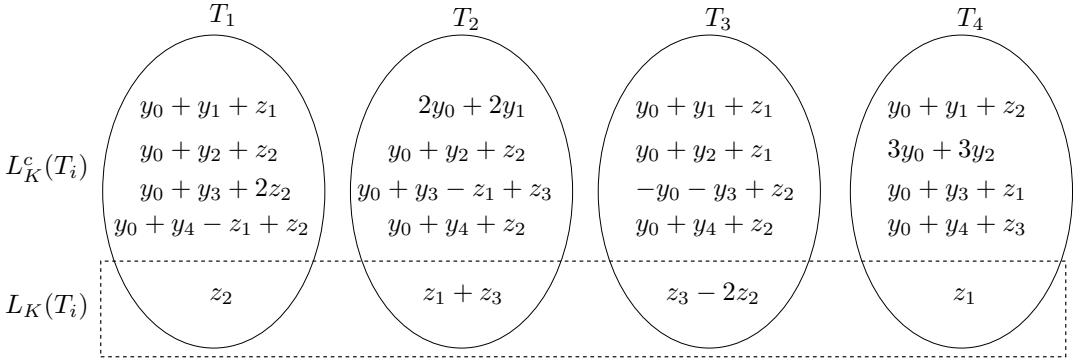


Fig. 2

3. SOME ALGEBRAIC LEMMAS

These are some algebraic claims that will be used throughout the various sections. They are proven through some standard algebraic degree arguments. The proofs are given in Section 8.

We remind the reader that an ideal I of $\mathcal{R}$ with generators $f_i, i \in [m]$, is the set $\{\sum_{i\in[m]} q_i f_i | q_i\text{'s} \in \mathcal{R}\}$. It is denoted by $\langle f_1, \ldots, f_m\rangle$. For any $f \in \mathcal{R}$, the following notations mean the same: $f \equiv 0 (\text{mod } I)$, $f \equiv 0 (\text{mod } f_1, \ldots, f_m)$, and $f \in I$.

An $f \in \mathcal{R}$ is called a *zerodivisor* of an ideal I (or mod I) if $f \notin I$ and there exists a $g \in \mathcal{R} \setminus I$ such that $fg \in I$. Let $u, v \in \mathcal{R}$. It is easy to see that if u is nonzero mod I and is a *non*-zerodivisor mod I then: $uv \in I$ iff $v \in I$. This can be seen as some sort of a "cancellation rule" for non-zerodivisors. We show such a cancellation rule in the case of special ideals arising in $\Sigma\Pi\Sigma$ circuits.

Definition 3.1 (*Radical-span*). Let $S := \{f_1, \ldots, f_m\}$ be multiplication terms generating an ideal I. Define the linear space $\text{radsp}(S) := \text{sp}(L(f_1) \cup \ldots \cup L(f_m))$.

When the set of generators S are clear from the context we will also use the notation $\text{radsp}(I)$. Similarly, $\text{radsp}(I, f)$ is shorthand for $\text{radsp}(S \cup \{f\})$.

Remark. Radical-span is motivated by the *radical* of an ideal but it is not quite that, for example, $\text{radical}(x_1^2, x_1x_2) = \langle x_1\rangle$ but $\text{radsp}(x_1^2, x_1x_2) = \text{sp}(x_1, x_2)$. It is easy to see that the ideal generated by radsp always contains the radical ideal.

LEMMA 3.2 (NON-ZERODIVISOR). *Let $f_1, \ldots, f_m$ be multiplication terms generating an ideal I, let $\ell \in L(\mathcal{R})$ and $g \in \mathcal{R}$. If $\ell \notin$ radsp(I) then: $\ell g \in I$ iff $g \in I$.*

All the ideals arising in this work are *homogeneous*, i.e. their generators are homogeneous polynomials. These ideals have some nice properties, as shown below. Degree $\deg(\cdot)$ refers to the total degree unless there is a subscript specifying the variable.

LEMMA 3.3 (HOMOGENEOUS IDEALS). *Say, $f_1, \ldots, f_m, g$ are homogeneous polynomials in $\mathcal{R}$. Then,*

1) If $\deg(g) < \deg(f_m)$ then: $g \in \langle f_1, \ldots, f_m\rangle$ iff $g \in \langle f_1, \ldots, f_{m-1}\rangle$.

2) If $\deg(g) = \deg(f_m)$ then: $g \in \langle f_1, \ldots, f_m\rangle$ iff $\exists a \in \mathbb{F}$, $(g + af_m) \in \langle f_1, \ldots, f_{m-1}\rangle$.

We give an important lemma about matchings. Intuitively, a matching represents linear relationships between forms in two multiplication terms. Algebraically, we will often encounter multiplication terms that are similar modulo an ideal. We show that these are basically equivalent.

Definition 3.4 ($L_U(\cdot), L_U^c(\cdot)$). For a multiplication term f and a subspace $U \subseteq L(\mathcal{R})$ define $L_U(f) := L(f) \cap U$ and $L_U^c(f) := L(f) \setminus U$.

LEMMA 3.5 (FROM CONGRUENCE TO MATCHING). *Let I be an ideal generated by multiplication terms $\{f_1, \ldots, f_m\}$ and define $U :=$ radsp(I). Let g, h be multiplication terms such that $g \equiv h \not\equiv 0 \pmod I$. Then there is a U-matching between $L_U(g), L_U(h)$ and one between $L_U^c(g), L_U^c(h)$.*

4. MATCHING THE TERMS IN AN IDENTITY: CONSTRUCTION OF MAT-NUCLEUS

4.1. Chinese Remaindering for Multiplication Terms

Traditionally, Chinese remaindering states: If two coprime polynomials (resp. integers) f, g divide a polynomial (resp. integer) h, then fg divides h. The key tool in constructing mat-nucleus is a version of Chinese remaindering specialized for multiplication terms but generalized to ideals. Similar methods appeared first in [Kayal and Saxena 2007] but we make this more formal and give a simpler proof. In particular, we avoid the use of local rings and Hensel lifting. We state our Chinese remaindering for multiplication terms as a neat *ideal decomposition* statement.

THEOREM 4.1 (IDEAL CHINESE REMAINDERING). *Let $f_1, \ldots, f_m, p, f, g$ be multiplication terms. Define the ideal $I := \langle f_1, \ldots, f_m \rangle$. Assume $L(p) \subseteq$ radsp(I) while, $L(f) \cap$ radsp$(I) = \emptyset$ and $L(g) \cap$ radsp$(I, f) = \emptyset$. Then,*

$$\langle I, pfg \rangle = \langle I, p \rangle \cap \langle I, f \rangle \cap \langle I, g \rangle.$$

PROOF. If h is a polynomial in $\langle I, pfg \rangle$ then clearly it is in each of the ideals $\langle I, p \rangle$, $\langle I, f \rangle$ and $\langle I, g \rangle$.

Suppose h is a polynomial in $\langle I, p \rangle \cap \langle I, f \rangle \cap \langle I, g \rangle$. There exist $i_1, i_2, i_3 \in I$ and $a, b, c \in \mathcal{R}$ such that,

$$h = i_1 + ap = i_2 + bf = i_3 + cg.$$

The second equation gives $bf \in \langle I, p \rangle$. Since radsp$(I, p) =$ radsp(I), we get $L(f) \cap$ radsp$(I, p) = \emptyset$. We can express $f = \prod_{i=1}^{j} \ell_i$, for $\ell_i \in \mathcal{R}$. We have $(b \prod_{i=1}^{j-1} \ell_i)\ell_j \in \langle I, p \rangle$, and $\ell_j \notin$ radsp(I, p). By Lemma 3.2, $b \prod_{i=1}^{j-1} \ell_i \in \langle I, p \rangle$. We repeat this argument for all ℓ_is and deduce that $b \in \langle I, p \rangle$. Hence $bf \in \langle I, p \rangle f \subseteq \langle I, pf \rangle$, and $h = i_2 + bf \in \langle I, pf \rangle$. This ensures the existence of $i_2' \in I$ and a polynomial b' such that,

$$h = i_2' + b'pf = i_3 + cg.$$

We repeat the argument for $cg \in \langle I, pf \rangle$. Since $L(g) \cap$ radsp$(I, pf) = L(g) \cap$ radsp$(I, f) = \emptyset$, repeated applications of Lemma 3.2 give us $c \in \langle I, pf \rangle$. Hence $cg \in \langle I, pf \rangle g \subseteq \langle I, pfg \rangle$ and $h = i_3 + cg \in \langle I, pfg \rangle$. This finishes the proof. □

We now come to one of the most important definitions in this paper. We break up a multiplication term into "nodes" with respect to an ideal.

Definition 4.2 (*Nodes*). Let f be a multiplication term and let I be an ideal generated by some multiplication terms. Since the relation "similarity mod radsp(I)" is an equivalence relation on $L(\mathcal{R})$, it partitions the list $L(f)$ into equivalence classes.

[rep$_I(f)$] For each such class pick a representative ℓ_i and define rep$_I(f) := \{\ell_1, \ldots, \ell_r\}$. (Note that form 0 can also appear in this set, it represents the class $L(f) \cap$ radsp(I).) This definition is not unique, but it is not an issue. We can set this arbitrarily.

[nod$_I(f)$] For each $\ell_i \in$ rep$_I(f)$, we multiply the forms in f that are similar to ℓ_i mod radsp(I). We define *nodes of f mod I* as the set of polynomials nod$_I(f) :=$

$\{M(L(f) \cap (\mathbb{F}^*\ell + \text{radsp}(I))) \mid \ell \in \text{rep}_I(f)\}$. (Remark: When $I = \{0\}$, nodes of f are just the coprime powers-of-forms dividing f.)

[...wrt a subspace] Let K be a linear subspace of $L(\mathcal{R})$. Clearly, the relation "similarity mod K" is an equivalence relation on $L(\mathcal{R})$. It will be convenient for us to also use notations $\text{rep}_K(f)$ and $\text{nod}_K(f)$. They are defined by replacing $\text{radsp}(I)$ in the above definitions by K.

Observe that the product of polynomials in $\text{nod}_I(f)$ just gives f. Also, modulo $\text{radsp}(I)$, each node is just a form-power ℓ^r. In other words, modulo $\text{radsp}(I)$, a node is rank-one term. The choice of the word "node" might seem a bit mysterious, but we will eventually construct paths through these. To pictorially see what is going on, think of each term T_i as a set of its constituent nodes.

We prove some consequences of the ideal Chinese remaindering theorem that will be very helpful in both Steps 1 and 2.

THEOREM 4.3. *Let I be an ideal and f a multiplication term. Let the set $nod_I(f)$ be $\{g_1, \ldots, g_r\}$. Then*

$$\langle I, f\rangle = \bigcap_{i\in[r]} \langle I, g_i\rangle.$$

PROOF. We have $f = \prod_{i\in[r]} g_i$. Let the corresponding representatives $\text{rep}_I(f) = \{\ell_1, \ldots, \ell_r\}$. If $r = 1$, then the theorem is trivially true. So assume $r \geq 2$. If $L(f)$ has a form in $\text{radsp}(I)$, then assume wlog that ℓ_1 is the representative of the class $L(f) \cap \text{radsp}(I)$. Define $G_i := \prod_{i<j\leq r} g_i$, for all $i \in [r-1]$.

We claim that for all $i \in [r-1]$, $L(G_i) \cap \text{radsp}(I, g_i) = \emptyset$. Let us complete the proof, given this statement. Start with representing $f = g_1G_1$. By Theorem 4.1, $\langle I, f\rangle = \langle I, g_1\rangle \cap \langle I, G_1\rangle$. Now, we write $G_1 = g_2G_2$, and again apply Theorem 4.1 to get $\langle I, G_1\rangle = \langle I, g_2\rangle \cap \langle I, G_2\rangle$. By repeated applications of Theorem 4.1, we finally prove that $\langle I, f\rangle = \bigcap_{i\in[r]}\langle I, g_i\rangle$.

Now we show that $L(G_i) \cap \text{radsp}(I, g_i) = \emptyset$. Since ℓ_1 is the representative of the class $L(f) \cap \text{radsp}(I)$, no form in $L(G_i)$ (for any i) can be in $\text{radsp}(I)$. This means that if $L(G_i) \cap \text{radsp}(I, g_i)$ has some ℓ, then $\ell \in (\mathbb{F}^*\ell_i + \text{radsp}(I))$. But this contradicts $\ell_{i+1}, \ldots, \ell_r$ not being similar to ℓ_i mod $\text{radsp}(I)$. □

We state a more useful corollary of this theorem.

COROLLARY 4.4. *Let $h \in \mathcal{R}$, f be a multiplication term, and let I be an ideal generated by some multiplication terms. Then, $h \notin \langle I, f\rangle$ iff $\exists g \in nod_I(f)$ such that $h \notin \langle I, g\rangle$.*

4.2. Applying Chinese Remaindering to $\Sigma\Pi\Sigma$ Circuits

We showed the effect of ideal Chinese remaindering on a single multiplication term f in Theorem 4.3. Now we show the effect on a *tuple* of sub-products, for example, appearing in a $\Sigma\Pi\Sigma$ circuit. Towards that we consider a tuple of iteratively defined ideals and nodes; this tuple of nodes we call a *path*.

Definition 4.5 (*Paths*). Let I be an ideal generated by some multiplication terms. Let $S = \{s_1 < \ldots < s_k\}$ be an ordered index set. Let $D = \sum_{s\in S} T_s$ be a $\Sigma\Pi\Sigma(k, d)$ circuit. Let v_s be a *sub-term* of T_s (i.e. $L(v_s) \subseteq L(T_s)$), for all $s \in S$. We call the tuple $(I, v_{s_1}, \ldots, v_{s_k})$ a *path of D mod I* if, for all $i \in [k]$, $v_{s_i} \in \text{nod}_{\langle I, v_{s_1}, \ldots, v_{s_{i-1}}\rangle}(T_{s_i})$. The *length* is the number of nodes in the path (here k).

To understand paths, let us intuitively explain what we want. Consider the special setting where each of the v_s are single forms and $I = 0$. Ignoring the use of nodes, the

tuple $(v_{s_1}, v_{s_2}, \ldots)$ form a *path* when (for all i) v_{s_i} is not contained in $\text{sp}(v_{s_1}, \ldots, v_{s_{i-1}})$. So, they form a "chain of linear independence".

One of the main technical challenges in computing rank bounds (and depth-3 PIT analysis in general) is that notions of linear independence do not suffice. We have to move to the language of ideals. Our definition of paths is in some sense a generalization of this chain of linear independence. We start with some ideal I. This ideal splits the forms in T_{s_1} into nodes, by looking at forms modulo $\text{radsp}(I)$. We pick some node v_{s_1}. Now, T_{s_2} is split into nodes by ideal $\langle I, v_{s_1} \rangle$. We pick a node v_{s_2}, and continue in this fashion to get a path. This leads to a path of nodes with some sort of polynomial independence.

We have defined path $\overline{p}$ as a tuple but, for convenience, we will sometimes treat it as a *set* of multiplication terms, eg. when operated upon by $\text{sp}(\cdot)$, $\langle \cdot \rangle$, $\text{radsp}(\cdot)$, etc. Abusing notation, when we say $(\text{mod } \overline{p})$, we mean modulo the ideal generated by I and the nodes in $\overline{p}$. Conventionally, when $k = 0$ the circuit C has just "one" gate: 0. In that case, the only path C has is (I), which is of length 0.

We also set some notation. For any subset $S \subseteq [k]$, the *sub-circuit* $C_S := \sum_{s \in S} T_s$. For an $i \in \{0, \ldots, k-1\}$, define $[i]^c := [k] \setminus [i]$. Conventionally, $[0] := \emptyset$ and $C_\emptyset := 0$.

Our following version of 'Chinese Remainder Theorem', directly inspired by [Kayal and Saxena 2007], states that if C is a nonzero $\Sigma\Pi\Sigma(k, d)$ circuit then there is a path that "certifies" that C is nonzero by reducing the whole circuit to a single nonzero multiplication term. Let us take a slight detour to understand the significance of this theorem. The goal of Chinese Remaindering can be thought of as specifying a set of "tests" that check if a circuit is not an identity. The guarantee is that for a non-identity, there is some test that correctly certifies this and identities fail all the tests. This leads to a viable algorithm if the set of tests is relatively small, and each test is easy to perform.

The theorem states that the paths of circuit give this list of tests. For a non-zero circuit, there exists a path $\overline{p}$ (for convenience, let the indices simply be $[i]$) with the following property. Consider the circuit modulo the corresponding ideal $\langle \overline{p} \rangle$. Then $C_{[i]^c} \pmod{\overline{p}}$ is the same (up to some constant factor) to a *single non-zero term* T_j ($j > i$).

This test can be algorithmically implemented because of the special nature of path ideals. Hence, for any path, we can efficiently check whether $C_{[i]^c}$ reduces to a single non-zero term modulo $\langle \overline{p} \rangle$. The total number of paths is at most d^k, so this gives a whitebox algorithm for depth-3 PIT. This is actually an overview of Kayal-Saxena algorithm [Kayal and Saxena 2007]. Thus, our theorem can be considered as a restatement of a key technical component of [Kayal and Saxena 2007].

Back to rank bounds. For our purposes, it is not the total number of paths, but the rank of the path that is important. A path consists of at most k nodes, so that rank of all the forms in the path is at most $k + \text{rk}(\text{radsp}(I))$. Hence, it is a *low-rank certificate for the nonzeroness of* C. This theorem is also central to later improvements for depth-3 PIT [Saxena and Seshadhri 2011b].

THEOREM 4.6 (CERTIFICATE FOR A NON-IDENTITY). *Let I be an ideal generated by some multiplication terms. Let $C = \sum_{i \in [k]} T_i$ be a $\Sigma\Pi\Sigma(k, d)$ circuit that is nonzero modulo I. Then $\exists i \in \{0, \ldots, k-1\}$ such that $C_{[i]}$ mod I has a path $\overline{p}$ satisfying: $C_{[i]^c} \equiv \alpha \cdot T_j \not\equiv 0$ (mod $\overline{p}$) for some $\alpha \in \mathbb{F}^*$ and $j \in [i]^c$.*

Suppose the reader has kept the mental picture of the terms as consisting of rank-one (modulo $\text{radsp}(I)$) nodes. A path $\overline{p}$ "kills" the terms that it passes through, and collapses the remaining circuit into a single term. This is very reminiscent of the poly-time algorithm of [Kayal and Saxena 2007]. Indeed, this theorem is a (shorter) proof

of the correctness of the algorithm. Why? Consider the path $\overline{p}$ given by the theorem when I is the zero ideal. The path $\overline{p}$ can be represented by a list of at most k 'forms' in $L(C)$. This path comes from some $C_{[i]}$, which means that $C_{[i]} = 0 (\text{mod } \overline{p})$. So, we get that $C \equiv \alpha \cdot T_{i+1} \not\equiv 0 (\text{mod } \overline{p})$. Since T_{i+1} is a product of linear forms, it is easy to algorithmically check if $C \equiv 0 (\text{mod } \overline{p})$. If C is identically zero, such a path cannot exist. Since there are at most d^k different paths, we can exhaustively check all of them. That yields an alternative view of the [Kayal and Saxena 2007] test.

Before giving the formal proof, we explain the main idea. Let us for convenience assume that no term is contained in I. We know that $C = T_1 + \cdots + T_k$ is not in I. We can argue that $C_{[1]^c} \notin \langle I, T_1 \rangle$ (otherwise we are done). By our Chinese Remainder Theorem (specifically, Corollary 4.4), there exists a node $g_1 \in \text{nod}_I(T_1)$ such that $C_{[1]^c} \notin \langle I, g_1 \rangle$. We then repeat the above argument considering $C_{[1]^c}$ and the ideal $\langle I, g_1 \rangle$. This yields the second node g_2. As we continue this argument, we will end up at the desired path. For technical reasons, we construct a more direct proof avoiding induction.

PROOF. Fix an $i \in \{0, \ldots, k-1\}$ and a path $\overline{p}$ of $C_{[i]}$ mod I such that:
(1) $C_{[i]^c} \notin \langle \overline{p} \rangle$.
(2) The cardinality of the set $J_{\overline{p}} := \{j \in [i]^c \mid T_j \notin \langle \overline{p} \rangle\}$ is the smallest possible (over all i).
(3) The index i is the largest one that attains the smallest cardinality described above.

Note that whenever the first condition is satisfied, $J_{\overline{p}} \neq \emptyset$. Hence for $i = 0$, $\overline{p} = (I)$, the first condition is satisfied and the corresponding $J_{\overline{p}} \neq \emptyset$. Thus, the desired i and $\overline{p}$ exist. We will argue that this path satisfies the conditions of the theorem.

Suppose $C_{[i]^c} \notin \langle \overline{p}, T_{i+1} \rangle$. By Corollary 4.4, there exists $v \in \text{nod}_{\langle \overline{p} \rangle}(T_{i+1})$ such that $C_{[i]^c} \notin \langle \overline{p}, v \rangle$. Note that $\overline{q} := (\overline{p}, v)$ forms a path of $C_{[i+1]} (\text{mod } I)$. Since $T_{i+1} \in \langle \overline{q} \rangle$, $C_{[i+1]^c} = C_{[i]^c} - T_{i+1} \notin \langle \overline{q} \rangle$. We also have $J_{\overline{q}} \subseteq J_{\overline{p}}$ (so $|J_{\overline{q}}| \leq |J_{\overline{p}}|$) simply because $\langle \overline{q} \rangle \supseteq \langle \overline{p} \rangle$. This violates either the second or third condition given above.

Thus, $C_{[i]^c} \in \langle \overline{p}, T_{i+1} \rangle$. The polynomials generating $\langle \overline{p}, T_{i+1} \rangle$ are all homogeneous, and so is $C_{[i]^c}$. By Lemma 3.3, there exists an $\alpha \in \mathbb{F}$ such that $(C_{[i]^c} - \alpha T_{i+1}) \in \langle \overline{p} \rangle$. Since $C_{[i]^c} \notin \langle \overline{p} \rangle$, the above equation can be rewritten as:

$$C_{[i]^c} \equiv \alpha T_{i+1} \not\equiv 0 \text{ (mod } \langle \overline{p} \rangle).$$

This completes the proof (α nonzero is implied). □

Remark. The above theorem only needs the non-zeroness of C mod I and has no simplicity or minimality requirements.

4.3. Using Minimality to get mat-nucleus

If we are given a circuit that is zero and minimal (not necessarily simple), then a repeated application of Theorem 4.6 gives us a space *mat-nucleus* that matches all the multiplication terms of C.

Theorem 2.2 (restated). Let $C = T_1 + \cdots + T_k$ be a $\Sigma\Pi\Sigma(k, d)$ circuit that is minimal and zero. Then there exists a linear subspace K of $L(\mathcal{R})$ such that:

1) $\text{rk}(K) < k^2$.
2) $\forall i \in [k]$, there is a K-matching π_i between T_1, T_i.

PROOF. We construct a set U, consisting of forms in $L(C)$, through an iterative process. Consider the relation in the set $[k]$, where i and j are related if T_i and T_j are U-matched. Note that this is an equivalence relation and hence partitions $[k]$. We will refer to this as the partition induced by U, denoted by $\mathcal{P}(U)$. The size of $\mathcal{P}(U)$ is the

number of sets in this partition. As we add forms to U, $\mathcal{P}(U)$ cannot increase. We will show how to add at most k forms to U to (strictly) decrease $\mathcal{P}(U)$. This suffices to complete the proof. When $U = \emptyset$, $|\mathcal{P}(U)|$ is at most k. When $\mathcal{P}(U)$ reaches 1, the linear span of U will be our desired K, and the rank of K will be smaller than k^2.

We now show how to decrease $|\mathcal{P}(U)|$ by adding at most k forms to U. Let S be some set in the current partition $\mathcal{P}(U)$. Since $C_S \neq 0$ (by minimality), we can apply Theorem 4.6 on C_S mod $\langle 0 \rangle$ to get a path $\overline{p}_S$ inside C_S mod $\langle 0 \rangle$ such that $\exists i \in S$, $C_S \equiv \alpha T_i \not\equiv 0 \pmod{\langle \overline{p}_S \rangle}$ for some $\alpha \in \mathbb{F}^*$.

Define $S^c := [k] \setminus S$. Now,

$$C \equiv C_{S^c} + \alpha T_i \equiv 0 \pmod{\langle \overline{p}_S \rangle}. \tag{1}$$

This means $C_{S^c} \notin \langle \overline{p}_S \rangle$ (otherwise $\alpha T_i \in \langle \overline{p}_S \rangle$, a contradiction). For convenience, let use relabel indices so that $S^c = [|S^c|]$. We can apply Theorem 4.6 on C_{S^c} mod $\langle \overline{p}_S \rangle$ to get a path $\overline{p}_{S^c}$ inside C_{S^c} mod $\langle \overline{p}_S \rangle$ and $j \in S^c$ such that, $C_{S^c} \equiv \beta T_j \not\equiv 0 (\mathrm{mod}\ \langle \overline{p}_{S^c} \rangle)$, for some $\beta \in \mathbb{F}^*$. Note that the ideal $\langle \overline{p}_{S^c} \rangle$ contains $\langle \overline{p}_S \rangle$, since the path $\overline{p}_{S^c}$ is constructed in C_{S^c} mod $\langle \overline{p}_S \rangle$. This allows us to rewrite Equation (1) as:

$$\alpha T_i \equiv -\beta T_j \not\equiv 0 \pmod{\langle \overline{p}_{S^c} \rangle}$$

Define $K' := \mathrm{radsp}(\overline{p}_{S^c})$. Observe that $\overline{p}_{S^c}$ is a path of some sub-circuit of C mod $\langle 0 \rangle$ and has length at most $|S| - 1 + |S^c| - 1 \leq k - 2$. Hence, $\mathrm{rk}(K') < (k-1)$. Also, by Lemma 3.5, the above congruence implies a K'-matching between T_i and T_j.

We add a basis of K' to U. Before adding K', $i \in S$ and $j \in S^c$ were not related, but they are related after this addition. Hence $\mathcal{P}(U)$ must have decreased in size. □

5. CERTIFICATE FOR LINEAR INDEPENDENCE: THE NUCLEUS

Suppose we have multiplication gates $T_1, \ldots, T_{k'}$ and a space K' of $L(\mathcal{R})$ such that T_1, T_i is K'-matched, for all $i \in [k']$. We show in this section that if $T_1, \ldots, T_{k'}$ are linearly independent (i.e. $\nexists\, \overline{\beta} \in \mathbb{F}^{k'} \setminus \{\overline{0}\}$ s.t. $\sum_{i \in [k']} \beta_i T_i = 0$) then K' can be extended to a linear space K of rank at most $(\mathrm{rk}(K') + k'^2)$ such that: $M(L_K(T_1)), \ldots, M(L_K(T_{k'}))$ are also linearly independent.

Theorem 2.4 (restated). Let $C = \sum_{i \in [k]} T_i$ be a minimal $\Sigma\Pi\Sigma(k, d)$ identity and let $\{T_i | i \in \mathcal{I}\}$ be a maximal set of linearly independent terms ($1 \leq k' := |\mathcal{I}| < k$). Then there exists a linear subspace K of $L(\mathcal{R})$ such that:

1) $\mathrm{rk}(K) < 2k^2$.
2) $\forall i \in [k]$, there is a K-matching π_i between T_1, T_i.
3) (Define $\forall i \in \mathcal{I}$, $K_i := M(L_K(T_i))$.) The terms $\{K_i | i \in \mathcal{I}\}$ are linearly independent.

PROOF. For convenience, and wlog, assume $\mathcal{I} = [k']$ and $k' \geqslant 2$. The proof is an iterative process with at most k'^2 iterations, and gradually builds the promised space K. Each iteration of the process maintains a space U of $L(\mathcal{R})$ that is intended to grow at each step and bring us closer to K. For convenience, define $U_i := M(L_U(T_i))$, for all $i \in [k']$. Also for each $i \in \{2, \ldots, k'\}$, define ideal $\mathcal{I}_i := \langle U_1, \ldots, U_{i-1} \rangle$.

The process has two nested iterations, or phrased differently, a double induction. We will call the outer "loop" a *phase*, and the inner loop a *round*. In each round the rank of U increases by at most 1, and the i-th phase has at most i rounds. At the end of the i-th phase ($i \geq 2$), we will ensure $T_i \notin \mathcal{I}_i$. Note that U_i is exactly the set of forms in T_i contained in $\mathrm{radsp}(\mathcal{I}_i)$. By Lemma 3.2, $T_i \notin \mathcal{I}_i$ is equivalent to $U_i \notin \mathcal{I}_i$.

In the first phase we set $U := K'$, where K' is the matching-nucleus obtained by applying Theorem 2.2 on C. This immediately gives us property (2) promised in the theorem statement, i.e. the matching property. Also, $\mathrm{rk}(U) < k^2$ at the end of the first phase.

Before we explain the remaining phases, let us explain why our construction suffices to complete the proof. We will add forms to U as the process continues, so the U-matching continues to hold. At the end, all the terms U_i have the same degree (since there exist U-matchings between all the terms).

Furthermore, we have guaranteed that $U_i \notin \mathcal{I}_i$ for all i. Since all U_i have the same degree, by repeated applications of Lemma 3.3, $U_i \notin \text{sp}(U_1, U_2, \ldots, U_{i-1})$. Therefore, the U_is are linearly independent, giving us property (3).

The rank bound holds since we only add a single form to U in each round (after the first phase), and there are at most k^2 rounds. Note that since we only grow U, the ideals $\mathcal{I}_i$ only become smaller. In other words, if we add more forms to U, any polynomial that was not in $\mathcal{I}_i$ will still not be in $\mathcal{I}_i$ (after the addition).

Back to the construction now. We explain the second phase. As T_1, T_2 are linearly independent, we get, by Lemma 3.3, that $T_2 \notin \langle T_1 \rangle$. By an application of Corollary 4.4, $\exists v \in \text{nod}_{\langle 0 \rangle}(T_1)$ such that $T_2 \notin \langle v \rangle$. We update $U \leftarrow (U + \text{radsp}(v))$. Note that after updating, $T_2 \notin \langle U_1 \rangle = \mathcal{I}_2$ (otherwise $T_2 \in \langle U_1 \rangle \subseteq \langle v \rangle$, since $v | U_1$).

Now, for the $i > 2$ phase. Inductively, we assume that $\forall r < i, T_r \notin \mathcal{I}_r$ (remember that all these ideals are wrt the *current* U). The phase consists of various rounds. At the end of the j-th round ($1 \leq j < i$), we just want to ensure $T_i \notin \langle U_1, \ldots, U_j, T_{j+1}, \cdots, T_{i-1} \rangle$. So we do nothing in the j-th round unless this is violated. What do we do when it is violated? The following is the technical core of the proof.

CLAIM 5.1. *Let $i > 2$ and $1 \leq j < i$. Suppose $\forall r < i$, $T_r \notin \langle U_1, \ldots, U_{r-1} \rangle$. Suppose $T_i \in \langle U_1, \cdots, U_j, T_{j+1}, \cdots, T_{i-1} \rangle$ but $T_i \notin \langle U_1, \cdots, U_{j-1}, T_j, \cdots, T_{i-1} \rangle$. There exists a $v \in \text{nod}_{\langle U_1, \cdots, U_{j-1} \rangle}(T_j)$ such that for the updated $U' \leftarrow (U + \text{radsp}(v))$ we have $T_i \notin \langle U'_1, \cdots, U'_j, T_{j+1}, \cdots, T_{i-1} \rangle$.*

Proof of Claim 5.1. Since $T_i \in \langle U_1, \cdots, U_j, T_{j+1}, \cdots, T_{i-1} \rangle$, by Lemma 3.3, we get $T_i + \sum_{r=j+1}^{i-1} \alpha_r T_r \in \langle U_1, \cdots, U_j \rangle$ for some α_r-s in $\mathbb{F}$. Suppose there are two distinct choices for α_r-s (we will call them α_r and α'_r). Then,

$$\left(T_i + \sum_{r=j+1}^{i-1} \alpha_r T_r \right), \left(T_i + \sum_{r=j+1}^{i-1} \alpha'_r T_r \right) \in \langle U_1, \cdots, U_j \rangle.$$

Subtracting, we get $\sum_{r=j+1}^{i-1} (\alpha - \alpha'_r) T_r \in \langle U_1, \cdots, U_j \rangle$. Let s be the largest index such that $\alpha_s - \alpha'_s \neq 0$. (By the distinctness of the sequences, such an index exists.) We get that $T_s \in \langle U_1, \cdots, U_j, T_{j+1}, \cdots, T_{s-1} \rangle \subseteq \langle U_1, \cdots, U_{s-1} \rangle$. Since $s \leq i-1$, this contradicts the hypothesis. Hence, the sequence $\{\alpha_r\}$ is unique.

The claim hypothesis says that $T_i \notin \langle U_1, \cdots, U_{j-1}, T_j, \cdots, T_{i-1} \rangle$. That implies $T_i + \sum_{r=j+1}^{i-1} \alpha_r T_r \notin \langle U_1, \cdots, U_{j-1}, T_j \rangle$. Thus, by Corollary 4.4, $\exists v \in \text{nod}_{\langle U_1, \cdots, U_{j-1} \rangle}(T_j)$ such that $T_i + \sum_{r=j+1}^{i-1} \alpha_r T_r \notin \langle U_1, \cdots, U_{j-1}, v \rangle$. Let us update U to $U' \leftarrow (U + \text{radsp}(v))$. (This updates U_r-s to U'_r-s.)

We now argue that $T_i \notin \langle U'_1, \ldots, U'_j, T_{j+1}, \cdots, T_{i-1} \rangle$. Suppose not. Then, by Lemma 3.3, for some sequence β_r, $T_i + \sum_{r=j+1}^{i-1} \beta_r T_r \in \langle U'_1, \ldots, U'_j \rangle \subseteq \langle U_1, \ldots, U_j \rangle$ (since for all r, $U_r | U'_r$). By the uniqueness of $\{\alpha_r\}$, we have $\beta_r = \alpha_r$, for all r. But that implies $T_i + \sum_{r=j+1}^{i-1} \alpha_r T_r \in \langle U'_1, \ldots, U'_j \rangle \subseteq \langle U_1, \cdots, U_{j-1}, v \rangle$. This is a contradiction and hence completes the proof. □

Let us look at the first round (i.e. $j = 1$). Suppose $T_i \notin \langle U_1, T_2, \cdots, T_{i-1} \rangle$. Then, we move directly to the second round, since we have already satisfied the round in-

variant. Otherwise, $T_i \in \langle U_1, T_2, \cdots, T_{i-1}\rangle$. Furthermore, by linear independence of $T_1, \ldots, T_i$ and Lemma 3.3, we have $T_i \notin \langle T_1, \cdots, T_{i-1}\rangle$, so we can invoke Claim 5.1 to get a $v \in \mathrm{nod}_{\langle 0\rangle}(T_1)$. This allows us to update $U \leftarrow (U + \mathrm{radsp}(v))$ such that $T_i \notin \langle U_1, T_2, \cdots, T_{i-1}\rangle$.

Now for the induction step. We assume that, by the end of the $(j-1)$th round, $T_i \notin \langle U_1, \cdots, U_{j-1}, T_j, \cdots, T_{i-1}\rangle$. For the j-th round, either we would have to do nothing or have to apply Claim 5.1 and update U. In either case, $\mathrm{rk}(U)$ increases by at most 1. At the end of the round, $T_i \notin \langle U_1, \cdots, U_j, T_{j+1}, \cdots, T_{i-1}\rangle$.

This continues till $j = i-1$. We finally have $T_i \notin \langle U_1, \cdots, U_{i-1}\rangle = \mathcal{I}_i$, giving us the required invariant for the i-th phase. This completes the proof. □

The following lemma proves the existence of the nucleus identity. The proof is given in Section 8.

LEMMA 5.2 (NUCLEUS IDENTITY). *Suppose $C = \sum_i T_i$ is a $\Sigma\Pi\Sigma(k,d)$ identity and K is a subspace of $L(\mathcal{R})$ such that T_1, T_i are K-matched, for all $i \in [k]$. Then the terms $M(L_K(T_i))$, for $i \in [k]$, are all of the same degree, say d', and form a $\Sigma\Pi\Sigma(k,d')$ identity $\sum_{i\in[k]} \alpha_i M(L_K(T_i))$, for some $\alpha_i \in \mathbb{F}^*$.*

6. INVOKING SYLVESTER-GALLAI THEOREMS: THE FINAL RANK BOUND

In this section we will bound the non-nucleus rank of a simple, minimal $\Sigma\Pi\Sigma(k,d)$, independent-fanin k', identity C by $(k-k')\cdot \mathrm{SG}_{k'}(\mathbb{F},d)$. That proves Theorem 2.11. We begin by dealing with strongly minimal case, which is really the hard part. The extension to simple, minimal identities follows with a little work. We will begin with some preliminaries definition. Then, we will give a high level picture of the overall strategy. The formal proof will follow, after which we show how to generalize to minimal identities.

Recall that if $C := \sum_{i\in[k]} T_i$ is a strongly minimal $\Sigma\Pi\Sigma(k,d)$ identity then $T_1, \ldots, T_{k-1}$ are linearly independent polynomials. Our aim is to bound the non-nucleus rank of such a simple C by $\mathrm{SG}_{k-1}(\mathbb{F},d)$, finishing the proof of Theorem 2.9.

6.1. Preliminaries

Fix K as the nucleus of C given by Theorem 2.4 with $\mathcal{I} = [k-1]$. There are two important properties of this nucleus that we restate (and elaborate upon) for emphasis.

The first is the *matching property*. For any $i \in [k]$, $L^c_K(T_1)$ $(= L(T_1)\setminus K)$ is K-matched to $L^c_K(T_i)$ $(= L(T_i)\setminus K)$. In other words for any $\ell \in L^c_K(T_1)$, the degrees of $M(L^c_K(T_1) \cap (\mathbb{F}^*\ell + K))$ and $M(L^c_K(T_i) \cap (\mathbb{F}^*\ell + K))$ are equal (these are polynomials in $\mathrm{nod}_K(T_1)$ and $\mathrm{nod}_K(T_i)$ respectively). This observation motivates the following definition.

Definition 6.1 (*Family*). Let C be a $\Sigma\Pi\Sigma(k,d)$ identity and K be its nucleus. For $\ell \in L^c_K(C)$, the *family of* ℓ is defined to be the list, $\mathrm{fam}(\ell) := \{M(L^c_K(T_i)\cap(\mathbb{F}^*\ell+K)) \mid i \in [k]\}$. Note that $\mathrm{fam}(\ell)$ is a multiset of size exactly k, having equal degree polynomials corresponding to each term T_i, we fix this ordering on the list (i.e. i-th element in $\mathrm{fam}(\ell)$ corresponds & divides the multiplication term T_i).

Verify that any two forms in $L^c_K(C)$ that are "similar mod K" have the same families.

[Partition, Class, Split & Preserve] Let us focus on a list $\mathrm{fam}(\ell)$. The equivalence relation of similarity (i.e. mod $\langle 0\rangle$) on $\mathrm{fam}(\ell)$, induces a *partition* of $[k]$ (i.e. if $f_i, f_j \in \mathrm{fam}(\ell)$ are similar then place i and j in the same partition-class). Denote this partition induced on $[k]$, by $\mathrm{Part}(\ell)$. Observe that $\mathrm{Part}(\ell)$ must contain at least 2 classes (otherwise simplicity of C is violated).

Each set in this partition is called a *class*, and we naturally have a class $\mathrm{cl}(f)$ associated with each member of $f \in \mathrm{fam}(\ell)$.

We say that Part(ℓ) *splits* a subset $S \subseteq [k]$ if there is some class $X \in$ Part(ℓ) such that $X \cap S \neq \emptyset, S$. Otherwise, we say that Part(ℓ) *preserves* S. Note that a singleton is always preserved.

For classes $A_1 \in$ Part(ℓ_1) and $A_2 \in$ Part(ℓ_2), the *complement* $\overline{A_1 \cup A_2}$ is just the set $[k] \setminus (A_1 \cup A_2)$. We will be later interested in the properties of this complement set wrt the two partitions.

The second property of the nucleus, the *linear independence*, will be used via the following claim. We define $K_i = M(L_K(T_i))$, for all $i \in [k]$, and by Lemma 5.2: $\sum_{i\in[k]} \alpha_i K_i = 0$ for some α_i's $\in \mathbb{F}^*$. The following holds quite directly from the nucleus properties.

CLAIM 6.2. *Suppose C is strongly minimal. For $1 < r < k$, let $\{s_1, \cdots, s_r\}$ be a subset $S \subsetneq [k]$. Then $K_{s_r} \notin \langle K_{s_1}, \cdots, K_{s_{r-1}}\rangle$.*

PROOF. Wlog assume $s_1 < s_2 < \cdots < s_r$. If $s_r < k$, then this just holds from the linear independence of $\{K_1, \ldots, K_{k-1}\}$ and Lemma 3.3. So, we can assume $s_r = k$ and $K_k \in \langle K_{s_1}, \cdots, K_{s_{r-1}}\rangle$. By Lemma 3.3, this means $K_k = \sum_{i\in[r-1]} \beta_{s_i} K_{s_i}$ for some β's $\in \mathbb{F}$. The nucleus identity gives us $K_k = -\sum_{i\in[k-1]} \frac{\alpha_i}{\alpha_k} K_i = \sum_{i\in[r-1]} \beta_{s_i} K_{s_i}$. Since $r < k$, this implies that for some γ's in $\mathbb{F}$, not all zero, $\sum_{i\in[k-1]} \gamma_i K_i = 0$. This contradicts the linear independence of $\{K_1, \ldots, K_{k-1}\}$, finishing the proof. □

Before applying Sylvester-Gallai-type theorems (i.e. the SG_{k-1} operator) we emphasize that, as discussed in Section 2.4, we fix a linear form $y_0 \in L(\mathcal{R})^*$ and a subspace U of $L(\mathcal{R})$ such that $L(\mathcal{R}) = \mathbb{F}y_0 \oplus U \oplus K$ and every form in $L^c_K(C)$ is monic wrt y_0. Thus, for every $\ell \in L^c_K(C)$ there exists a unique way to express $\ell = \alpha y_0 + u + v$ ($\alpha \in \mathbb{F}^*$, $u \in U$ and $v \in K$). This is formalized in the following lemma, which is proven in Section 8. We can now define the truncation operator: trun(ℓ) = $y_0 + \alpha^{-1}u$.

LEMMA 6.3 (MONIC FORMS). *Let $|\mathbb{F}| > d$ and C be a $\Sigma\Pi\Sigma(k, d)$ identity, over $\mathbb{F}$, with nucleus K. Let $y_0 \in L(\mathcal{R})^*$ and U be a subspace of $L(\mathcal{R})$ such that $L(\mathcal{R}) = \mathbb{F}y_0 \oplus U \oplus K$. Then there exists an invertible linear transformation $\tau : L(\mathcal{R}) \to L(\mathcal{R})$ that fixes K and satisfies:*

1) *$\tau(C)$ is also a $\Sigma\Pi\Sigma(k, d)$ identity with nucleus K and the same simplicity, minimality properties.*
2) *Every form in $L^c_K(\tau(C)) = \tau(L^c_K(C))$ is monic wrt y_0.*

6.2. Proof strategy

We describe the overall proof strategy for Theorem 2.9. The formal proof actually goes through a (somewhat) convoluted contradiction. So we give an intuitive explanation of the ideas with the caveat that the actual proof may not follow the same argument.

As we mentioned earlier, we wish to bound the non-nucleus rank of C by $\text{SG}_{k-1}(\mathbb{F}, d)$. All the non-nucleus parts of the individual terms are matched modulo K, by Theorem 2.4. Hence, it suffices to bound the rank of trun($L^c_K(T_1)$), the truncated forms of the non-nucleus part of T_1. We will show that this set is SG_{k-1}-closed. Consider some linearly independent forms $\{\ell'_1, \ldots, \ell'_{k-1}\}$ in trun($L^c_K(T_1)$). By the matching property, there exists $\ell_i \in L(T_i)$ such that trun(ℓ_i) $= \ell'_i$. Let us look at C modulo $I := \langle \ell_1, \ldots, \ell_{k-1}\rangle$. This equals $T_k (\text{mod } I)$. This, with the zeroness of C, will imply that there exists $\ell_k \in L(T_k)$ such that $\{\ell_i | i \in [k]\}$ are linearly dependent. Suppose that ℓ_k is non-similar to ℓ_i, for all $i < k$. So $\ell'_k :=$ trun(ℓ_k) cannot be equal to ℓ'_i for any other i. By the matching property, there must be an $\ell''_k \in L(T_1)$ such that trun(ℓ''_k) $= \ell'_k$. Hence, we have shown that within trun($L^c_K(T_1)$), there exists a non-trivial linear combination of $\{\ell'_1, \ldots, \ell'_{k-1}\}$. If this would happen for *all* such sets, then we would prove that trun($L^c_K(T_1)$) is SG_{k-1}-closed.

But we were very lucky that ℓ_k was non-similar to the other ℓ_i's. So let us further generalize the argument. Consider some linearly independent forms $\{\ell'_1, \ldots, \ell'_r\}$ (for $r < k$) in $\text{trun}(L^c_K(T_1))$. Suppose there exist forms $\{\ell_1, \ldots, \ell_r\}$ with the following properties. First define for all $i \leq r$, $A_i := \{j | \ell_i \text{ divides } T_j\}$. For all $i \leq r$, $\text{trun}(\ell_i) = \ell'_i$. Furthermore, $\overline{\bigcup_{i \leq r} A_i}$ is a singleton, say $\{s\}$. Again, set $I := \langle \ell_1, \ldots, \ell_r \rangle$. If we look at $C \pmod I$, then every $T_i, i \in \bigcup_{i \leq r} A_i$ is trivially "killed". We are left with $T_s = 0 \pmod I$, and some form $\ell \in L(T_s)$ such that $\ell \in I$. Since ℓ *cannot* be similar to any one of the ℓ_i's, we will get a non-trivial linear dependence in $\text{trun}(L^c_K(T_1))$.

But luck has consistenly been with us, since we get $\overline{\bigcup_{i \leq r} A_i}$ to be a singleton. Ideally, we would like a very controlled way of killing terms. If we can add generators to our ideal I such that each generator only kills one term, then we can make this argument work. But we also want the generator to be of low rank, so that we can get linear dependencies. One of the main hurdles with this approach is that there might be subcircuits of the form $x^3y + x^2y^2 + xy^3$, where it is not possible to kill a single term using a single form. It is possible to go modulo (say) x^3 and selectively kill the first term. Since, in the end, we are only after linear dependencies modulo K, we might even go modulo polynomials that are form-powers modulo K. These are exactly members of a family.

So let us start with a set of forms $\{\ell_1, \ell_2, \ldots, \ell_r\}$ (for $r < k$) in $L^c_K(T_1)$. We wish to find some $\ell_{r+1} \in L^c_K(T_1)$ that is a non-trivial linear combination of these modulo K. Let us select polynomials $p_i \in \text{fam}(\ell_i)$, and set $I = \langle p_1, \ldots, p_r \rangle$ and $A_i = \{j | p_i \text{ divides } T_j\}$. Let $S = \overline{\bigcup_{i \leq r} A_i}$ and so we get $C_S = 0 \pmod I$. For each $i \leq r$ and $s \in S$, there is some member of $\text{fam}(\ell_i)$ dividing T_s. Suppose we were able to choose the p_i's such that for each i, the members of $\text{fam}(\ell_i)$ dividing each T_s is the *same* polynomial. Then we can divide by all these polynomials, and get $C'_S = 0 \pmod I$, where none of the linear forms involved are equal to any $\ell_i \pmod K$. So we are setting ourselves up to find a non-trivial linear dependency. But, what if S is not a singleton?

The linear independence properties of the nucleus portion, the K_i's, will save the day. Let $S = \{s_1, s_2, \ldots, s_a\}$. For each $s \in S$, we can selectively kill T_s by going modulo K_s. In other words, going modulo $\langle I, K_{s_1} \rangle$ will only kill terms T_j, where $j \in \bigcup_{i \leq r} A_i \cup \{s_1\}$. By going modulo $I' := \langle I, K_{s_1}, K_{s_2}, \ldots, K_{s_{a-1}} \rangle$, we will get $T_{s_a} = 0 \pmod{I'}$ without "trivially" killing T_{s_a}. So we can pull out a linear dependence modulo K involving some $\ell_{r+1} \in L^c_K(T_1)$ and the $\{\ell_1, \ell_2, \ldots, \ell_r\}$.

All of this hinged on the choice of the p_i's in the respective families so that we get the special property that members of $\text{fam}(\ell_i)$ dividing each T_s are the same. Amazingly, we can always choose the p_i's to ensure this. This we prove through a contradiction. If the rank of $\text{trun}(L^c_K(T_1))$ is too large, then (using the SG_{k-1} operator) we show that a family of $(k-1)$ partitions must have a peculiar property (Lemma 6.5). Lemma 6.7 shows that this property can never hold.

6.3. The actual proof

We first state a technical cancellation lemma, whose proof is in Section 8.

LEMMA 6.4 (CANCELLATION). *Let K be some subspace of $L(\mathcal{R})$ and let $\ell_1, \ldots, \ell_m \in L(\mathcal{R}) \setminus K$ be linearly independent modulo K. Let $f_1, \ldots, f_m$ be multiplication terms similar to powers of $\ell_1, \ldots, \ell_m$ respectively modulo K (i.e. each form in f_i is in $(\mathbb{F}^* \ell_i + K)$). Let $\ell \in L(\mathcal{R})^*$ such that for some $s \in [m]$, $\ell \in \mathbb{F}\ell_s + K$. Then, for any polynomial $f \in \mathcal{R}$,*

$$\ell f \in \langle f_1, \ldots, f_m \rangle \text{ iff } f \in \langle f_1, \ldots, \frac{f_s}{\gcd(f_s, \ell)}, \ldots, f_m \rangle.$$

LEMMA 6.5 (PARTITIONS FROM SG_{k-1}-TUPLE). *Suppose* $\mathrm{rk}(\mathrm{trun}(L_K^c(T_1))) > \mathrm{SG}_{k-1}(\mathbb{F}, d)$. *There exists a set* $\{\ell_1, \ell_2, \ldots, \ell_{k-1}\}$ *of* $k-1$ *forms in* $L_K^c(T_1)$ *with the following property. For any non-empty subset* $\mathcal{I} \subseteq [k-1]$ *and any collection of sets* $\{A_i | i \in \mathcal{I}\}$ *where* $A_i \in \mathit{Part}(\ell_i)$, *the set* $S := \overline{\bigcup_{i\in\mathcal{I}} A_i}$ *is either empty or split by* $\mathit{Part}(\ell_c)$, *for* some $c \in \mathcal{I}$.

PROOF. Since $\mathrm{rk}(\mathrm{trun}(L_K^c(T_1))) > \mathrm{SG}_{k-1}(\mathbb{F}, d)$, we can apply the SG_{k-1} operator on this set. Let the output of $\mathrm{SG}_{k-1}(\mathrm{trun}(L_K^c(T_1)))$ be the set $\{\ell'_1, \ell'_2, \cdots, \ell'_{k-1}\}$. For all $i \in [k-1]$, let $\ell_i \in L_K^c(T_1)$ be a form satisfying $\mathrm{trun}(\ell_i) = \ell'_i$. We will prove that this is the desired set of forms. We show that for all choices of $\mathcal{I}$ and the sets A_i, if $S = \overline{\bigcup_{i\in\mathcal{I}} A_i} \neq \emptyset$, then S is split by some $\mathrm{Part}(\ell_c)$.

This is shown by contradiction. Suppose there is some choice of $\mathcal{I}$ and sets A_i, where $S \neq \emptyset$ and S is preserved by $\mathrm{Part}(\ell_i)$, for all $i \in \mathcal{I}$. For all $i \in \mathcal{I}$, there exists an $f_i \in \mathrm{fam}(\ell_i)$ such that $A_i = \mathrm{cl}(f_i)$. Similarly, for all $i \in \mathcal{I}$, there exists a $g_i \in \mathrm{fam}(\ell_i)$ such that $S \subseteq \mathrm{cl}(g_i)$. The sets A_i and S are disjoint, so the classes $\mathrm{cl}(f_i)$ and $\mathrm{cl}(g_i)$ are different. The polynomials f_i, g_i are not similar, for all $i \in \mathcal{I}$. Let $S = \{s_1, s_2, \ldots, s_r\}$. The meat of the proof is the following claim. Define $\hat{K}$ to be the set $\bigcup_{i\in\mathcal{I}}(\mathbb{F}^*\ell_i + K)$.

Claim: There exists a form $\ell \in L(T_{s_r})$ such that $\ell \in (\mathrm{sp}(\ell_i | i \in \mathcal{I}) + K) \setminus \hat{K}$.

Proof: Define ideal $I := \langle f_i | i \in \mathcal{I}\rangle$. Let us focus on the sub-circuit $C_S = \sum_{j\in S} T_j$. Since $C = 0$ and $S = \overline{\bigcup_{i\in\mathcal{I}} \mathrm{cl}(f_i)}$, we deduce $C_S \in I$ (as f_i "kills" the term T_r for all $r \in \mathrm{cl}(f_i)$). For all $i \in \mathcal{I}$, since $S \subseteq \mathrm{cl}(g_i)$, g_i divides T_j ($j \in S$). But all these g_i's are pairwise coprime, since they come from different families. Hence, $\prod_{i\in\mathcal{I}} g_i$ divides T_j, for all $j \in S$. The multiplication term $T'_j := T_j/(\prod_{i\in\mathcal{I}} g_i)$ has no form in $\hat{K}$. Rewrite,

$$C_S \;=\; \left(\prod_{i\in\mathcal{I}} g_i\right) \cdot \left(\sum_{j\in S} T'_j\right) \;\in\; \langle f_i \mid i \in \mathcal{I}\rangle.$$

By a repeated application of Lemma 6.4 on the above system, we get:

$$\sum_{j\in S} T'_j \in \langle f'_i \mid i \in \mathcal{I}\rangle =: I', \text{ where, } f'_i := \frac{f_i}{\gcd(f_i, g_i)}, \forall i \in \mathcal{I}. \tag{2}$$

Since f_i, g_i are not similar, f'_i has degree ≥ 1, for all $i \in \mathcal{I}$. Since we have only changed the non-nucleus part of T_j to get T'_j, we deduce $K_{s_i} | T'_{s_i}$, for all $i \in [r]$. Define the ideal $I'' := \langle I', K_{s_1}, \cdots, K_{s_{r-1}}\rangle$. By Equation (2), $T'_{s_r} \in I''$. We have $\mathrm{radsp}(I'') \subseteq \mathrm{sp}(\ell_i \mid i \in \mathcal{I}) + K$. Let us factor $T'_{s_r} = B_0 B_1$, where B_0 is the product of all forms in $\mathrm{radsp}(I'')$ and B_1 is the remaining product. Thus, $B_0 B_1 \in I''$. By Lemma 3.2, B_1 can be cancelled out and we get $B_0 \in I''$.

Suppose all forms of B_0 are in K, so $B_0 = K_{s_r}$. This means $K_{s_r} \in I''$ implying,

$$K_{s_r} \in \langle K_{s_1}, \cdots, K_{s_{r-1}}, \{f'_i \mid i \in \mathcal{I}\}\rangle. \tag{3}$$

Recall that each form in f'_i is similar to some form in $(\mathbb{F}^*\ell_i + K)$, for all $i \in \mathcal{I}$. Suppose form $(\beta_i \ell_i + u_i) | f'_i$, for all $i \in \mathcal{I}$, for some β's in $\mathbb{F}^*$ and u's in K. In Equation (3) make the *evaluation*: $\ell_i \leftarrow -\beta_i^{-1} u_i$, for all $i \in \mathcal{I}$. This is a valid evaluation since $\{\ell_i \mid i \in \mathcal{I}\}$ are linearly independent mod K, and values substituted are from K. Clearly, this evaluation leaves the linear subspaces K_s ($s \in S$) unchanged, but zeroes out f'_i. Thus, we get $K_{s_r} \in \langle K_{s_1}, \cdots, K_{s_{r-1}}\rangle$, contradicting Claim 6.2.

As a result, we have a form $\ell | B_0$ such that $\ell \notin K$. We have $\ell \in \mathrm{radsp}(I'') \subseteq \mathrm{sp}(\ell_i \mid i \in \mathcal{I}) + K$. Since T'_{s_r} has no form in $\hat{K}$, $\ell \notin \hat{K}$. □

By the matching property of the nucleus, this in turn gives us an $\ell \in L^c_K(T_1)$ such that: $\ell \in (\text{sp}(\ell_i | i \in \mathcal{I}) + K) \setminus \hat{K}$. This means that there exist constants β_i's in $\mathbb{F}$, not all zero, such that $\ell \in \sum_{i\in\mathcal{I}} \beta_i \ell_i + K$. Hence, $\text{trun}(\ell)$ is a linear combination of $\{\text{trun}(\ell_i) | i \in \mathcal{I}\}$. Because $\text{trun}(\ell) \notin \bigcup_{i\in\mathcal{I}}(\mathbb{F}^*\text{trun}(\ell_i))$, this must be a non-trivial combination. But this contradicts the fact that $\{\ell_1, \cdots, \ell_{k-1}\}$ were obtained from $\text{SG}_{k-1}(\text{trun}(L^c_K(T_1)))$.

This contradiction proves that S is split by $\text{Part}(\ell_i)$, for some $i \in \mathcal{I}$. □

We will prove that the conditions on partitions given by Lemma 6.5 cannot occur.. Since the proof is fairly involved, we present that in the next subsection. For now, we give the necessary definitions and claims and complete the rank bound proof for Theorem 2.9. We have a universe $\mathcal{U} := [k]$ of elements.

Definition 6.6 (*Unbroken chain*). A partition of $\mathcal{U}$ is *trivial* if it contains the single set $\mathcal{U}$.

Let $\mathfrak{P}$ be a collection of non-trivial partitions of $\mathcal{U}$ (here a collection refers to a *multiset*, i.e. $\mathfrak{P}$ can have partitions repeated). A *chain in* $\mathfrak{P}$ is a sequence of sets $A_1, A_2, \cdots, A_s$ (for some s) such that each set comes from a different element of $\mathfrak{P}$ (say $A_i \in \mathcal{P}_i \in \mathfrak{P}$).

The chain $A_1, A_2, \cdots, A_s$ is an *unbroken chain*, if $\overline{\bigcup_{i\in[s]} A_i}$ is non-empty and *preserved* in $\mathcal{P}_j$, for each $j \in [s]$. (Here, we use the natural extension of the previous definition of 'preserve' to all partitions.)

Note that if $\overline{\bigcup_{i\leq s} A_i}$ is a singleton then it is trivially preserved in any partition, therefore, such a chain would be unbroken. By Lemma 6.5, the collection $\{\text{Part}(\ell_i) | i \in [k-1]\}$ has no unbroken chain. We will show that this is impossible. The following combinatorial lemma implies Theorem 2.9.

LEMMA 6.7 (PARTITIONS HAVE UNBROKEN CHAIN). *Let $\mathfrak{P}$ be a collection of non-trivial partitions of $\mathcal{U}$. If $\mathfrak{P}$ contains at least $|\mathcal{U}| - 1$ partitions then $\mathfrak{P}$ contains an unbroken chain.*

We put it together to prove that the non-nucleus rank of a simple and strongly minimal $\Sigma\Pi\Sigma(k, d)$ identity over $\mathbb{F}$ is at most $\text{SG}_{k-1}(\mathbb{F}, d)$.

PROOF. (of Theorem 2.9) Let $C = \sum_{i\in[k]} T_i$ be a simple and strongly minimal $\Sigma\Pi\Sigma(k, d)$ identity over $\mathbb{F}$, and let K be the nucleus provided by Theorem 2.4. As $|\mathbb{F}| > d$ we can assume (wlog by Lemma 6.3) the existence of a truncation operator on $L^c_K(T_1)$. We will show that the rank of $\text{trun}(L^c_K(T_1))$ is at most $\text{SG}_{k-1}(\mathbb{F}, d)$. By the matching property of the nucleus, $\text{trun}(L^c_K(T_1))$ together with K span $L(C)$. Therefore, a non-nucleus rank bound of the former suffices to bound the non-nucleus rank of $L(C)$.

Assuming that the rank of $\text{trun}(L^c_K(T_1))$ is greater than $\text{SG}_{k-1}(\mathbb{F}, d)$, as in Lemma 6.5, we invoke $\text{SG}_{k-1}(\text{trun}(L^c_K(T_1)))$ to get $\{\ell_1, \cdots, \ell_{k-1}\}$ in $L^c_K(T_1)$. Associated with each of these, we have the partition $\text{Part}(\ell_i)$. There are $k-1$ partitions in the collection $\mathfrak{P} := \{\text{Part}(\ell_i) | i \in [k-1]\}$, which are all non-trivial by the simplicity of C. Lemma 6.7 tells us that $\mathfrak{P}$ has an unbroken chain, while Lemma 6.5 says that $\mathfrak{P}$ has none. This contradiction implies the rank of $\text{trun}(L^c_K(T_1))$ is at most $\text{SG}_{k-1}(\mathbb{F}, d)$, thus finishing the proof.

Our proof shows an even stronger property: $\text{trun}(L^c_K(T_1))$ is SG_{k-1}-closed. □

6.3.1. Proof of Lemma 6.7. Intuitively, when the partitions in $\mathfrak{P}$ have many classes then an unbroken chain should be easy to find. For example, when $(k-1)$ partitions in $\mathfrak{P}$ are all equal to $\{\{1\}, \ldots, \{k\}\}$ then there is an easy unbroken chain, namely $\{1\}, \ldots, \{k-1\}$. On the other hand, when the partitions in $\mathfrak{P}$ contain few classes then we can effectively

decrease the universe and apply induction. Most of this subsection would deal with the former case. Let us first define the splitting property.

Definition 6.8 (*Splitting property*). Let $\mathfrak{P}$ be a collection of partitions of $\mathcal{U}$. Suppose for all non-empty $S \subset \mathcal{U}$, S is split by at least $(|S| - 1)$ partitions of $\mathfrak{P}$. Then $\mathfrak{P}$ is said to have the *splitting property*.

LEMMA 6.9. *Let $\mathfrak{P}$ be a collection of at least $(k-1)$ non-trivial partitions of $[k]$. If $\mathfrak{P}$ has the splitting property then there is a chain $A_1, \cdots, A_{k-1}$ in $\mathfrak{P}$ such that $\overline{\bigcup_{i \leq k-1} A_i} = \{k\}$. (In particular, $\mathfrak{P}$ has an unbroken chain.)*

We defer its proof and, instead, first show why this lemma would suffice.

PROOF. (of Lemma 6.7) We will prove this by induction on the universe size k. For the base case, suppose $k = 3$ and $\mathfrak{P} = \{\mathcal{P}_1, \mathcal{P}_2, \ldots\}$. So we have at least two partitions. If any partition (say $\mathcal{P}_1$) contains exactly 2 sets, it must be a pair and a singleton (say $\mathcal{P}_1 = \{\{1,2\},\{3\}\}$). But then $\{1,2\}$ is itself an unbroken chain in $\mathfrak{P}$. So, all the partitions can be assumed to consist only of singletons. But then we can take the set, say, $\{1\}$ from $\mathcal{P}_1$ and, say, $\{2\}$ from $\mathcal{P}_2$ to get an unbroken chain.

Now for the induction step. Suppose $\mathfrak{P}$ has at least $(k-1)$ partitions. We assume that the claim is true for universes of size upto $(k-1)$. If $\mathfrak{P}$ has the splitting property, then we are done by Lemma 6.9. If not, then for some subset $S \subset \mathcal{U}$ of size at least 2, S is split in at most $(|S|-2)$ partitions. Let the collection of partitions in $\mathfrak{P}$ that preserve S be $\mathfrak{P}'$. So $\mathfrak{P}'$ contains at least $(k-1) - (|S|-2) = (k - |S| + 1)$ partitions. Merge the elements of S into a new element, to get a new universe $\mathcal{U}'$ of size $(k - |S| + 1)$. The partitions in $\mathfrak{P}'$ are valid partitions of $\mathcal{U}'$, and still maintain their structure. We now have a universe of size $k - |S| + 1 < k$, and at least $k - |S| + 1$ partitions. By the induction hypothesis, there is an unbroken chain in $\mathfrak{P}'$. Observe that it is (under the natural correspondence) still an unbroken chain in the original collection $\mathfrak{P}$, and we are done. □

PROOF. (of Lemma 6.9) We will find partitions $\mathcal{P}_1, \cdots, \mathcal{P}_{k-1}$ in $\mathfrak{P}$ such that $\mathcal{P}_i$ splits $\{i,k\}$, for all $i \in [k-1]$. Thus, there is a set $A_i \in \mathcal{P}_i$ that contains i but not k. Naturally, $\overline{\bigcup_{i \leq k} A_i} = \{k\}$.

This labelling is constructed through an iterative process that goes through phases. In the ith phase, we will find $\mathcal{P}_i$. At the beginning of this phase, we have already determined $\mathcal{P}_1, \cdots, \mathcal{P}_{i-1}$ with the desired property and the remaining pool $\mathfrak{P}'$ of unlabelled partitions. During this phase, we may label some partition from $\mathfrak{P}'$ as $\mathcal{P}_b$ (for some $b < i$) and move the "old" $\mathcal{P}_b$ to $\mathfrak{P}'$. The property that $\mathcal{P}_b$ splits $\{b,k\}$ is always maintained. We will repeatedly perform this swapping until we find an unlabelled partition that splits $\{i,k\}$. At this point, we label this $\mathcal{P}_i$ and end this phase.

The first phase is easy to understand. By the splitting property, there is some partition that splits $\{1,k\}$. We set this to $\mathcal{P}_1$ and end Phase 1.

For all the other phases, we have some auxiliary data that is maintained. In the ith phase, we maintain a partition (for convenience, we will call this a "division") of $[i-1]$, $E_1, \cdots, E_{i-1}$. Think of these as indices of the currently labelled partitions $\mathcal{P}_1, \mathcal{P}_2, \ldots, \mathcal{P}_{i-1}$, as well as elements in the universe $\mathcal{U}$. We set $E_0 = \{i,k\}$, which will be fixed throughout this phase, and set $E_{\leq j} = \bigcup_{0 \leq l \leq j} E_l$.

For each set E_j, we have a set of partitions $\mathcal{C}_j$ corresponding to this index set ($:= \{\mathcal{P}_b | b \in E_j\}$). We fix $\mathcal{C}_0 = \emptyset$. We get a similar set of partitions $\mathcal{C}_{\leq j} = \bigcup_{1 \leq l \leq j} \mathcal{C}_l$. Note the difference, because the indices for this start from 1. This is because there is no partition associated with elements i and k.

There is a deliberate (and probably unfortunate) overloading of indices here. We use 'i' to denote both the phase and an element of our universe. There is a clear link between the two since the ith phase is attempting to split $\{i, k\}$. For clarity, we will denote the phase number by $\underline{i}$.

The $\underline{i}$th phase will continually change this division by "promoting" elements. This means that an element in E_j will be put in E_{j+1}, and this is the only way in the which the division changes. At the beginning of the $\underline{i}$th phase, we initialize $E_0 = \{i, k\}$ and $E_1 = [i-1]$ (so naturally, all other E_j's are empty). We set $lim = 1$, which is the largest j such that E_j is non-empty. We define a recursive procedure *Update*($\underline{i}$), and prove certain properties about its behavior. These will suffice to complete the proof.

THE PROCEDURE *Update*($\underline{i}$)

(1) Check if there is a partition $\mathcal{P} \in \mathfrak{P}'$ that splits $E_0 = \{i, k\}$. If so, label $\mathcal{P}$ as $\mathcal{P}_i$, output success and terminate entire program.
(2) For all $1 \leq j \leq lim$ in increasing order
  — For all $c \in E_j$
    — If there is a partition $\mathcal{P} \in \mathfrak{P}'$ that splits $E_{\leq j-1} \cup \{c\}$
      — Label $\mathcal{P}$ as $\mathcal{P}_c$, and move the old $\mathcal{P}_c$ to $\mathfrak{P}'$. Move c from E_j to E_{j+1}. If E_{j+1} is now a singleton, add 1 to lim. Call *Update* recursively.
(3) Output failure and terminate entire program.

A few comments. When *Update*($\underline{i}$) outputs success, it has indeed found a partition that splits $\{i, k\}$ so we are truly done. Although the procedure makes recursive calls, it never returns to an older call. This is because it reaches the failure step and the whole program terminates. The only possibilities for this program are success, failure, or infinite running. Failure occurs when the 'if' condition never holds. So for $1 \leq j \leq lim$ in increasing order and all $c \in E_j$, no partition $\mathcal{P} \in \mathfrak{P}'$ splits $E_{\leq j-1} \cup \{c\}$.

We will argue that a call to *Update*($\underline{i}$) from the initialization always results in a success. The proof of this is broken down into some simpler claims.

CLAIM 6.10. *If, during a call to* Update($\underline{i}$)*, the procedure completes jth iteration in Step 2, then all partitions in $\mathfrak{P}'$ preserve $E_{\leq j}$. Hence, the partitions labelled $\mathcal{P}_c$ will always split $\{c, k\}$.*

Proof of Claim 6.10. The current run of *Update*($\underline{i}$) went through the jth iteration of Step 2 without making a recursive call. Hence, all partitions in $\mathfrak{P}'$ preserve $E_{\leq j-1} \cup \{c\}$, for all $c \in E_j$. Therefore, all partitions in $\mathfrak{P}'$ must preserve $E_{\leq j-1} \cup E_j = E_{\leq j}$.

Whenever a partition is labelled $\mathcal{P}_c$ (in say the jth iteration), $\mathcal{P}_c$ splits $E_{\leq j-1} \cup \{c\}$ but preserves $E_{\leq j-1}$. For any $j \geq 1$, $E_{\leq j-1} \supseteq E_0 = \{i, k\}$. Hence, $\mathcal{P}_c$ splits $\{c, k\}$. □

CLAIM 6.11. *At all times, for $l \geq 2$, $\mathcal{C}_l$ preserves $E_{\leq l-2}$.*

Proof of Claim 6.11. Initially, this is vacuously true, since E_l is empty for $l \geq 2$. We will show that this is maintained whenever the division sets E_j (and the labelled partitions) change. The promotion of an element by moving from E_j to E_{j+1} can only decrease $E_{\leq l}$ for all l. Hence, a labelled partition that originally preserved some $E_{\leq l}$ continues to do so. Consider the new partition that is swapped in to become $\mathcal{P}_c$, during iteration (say) j. This will be a part of $\mathcal{C}_{j+1}$, because c will be moved to E_{j+1}. If $j = 1$, then this partition $\mathcal{P}_c$ must preserve $\{i, k\} = E_{\leq 0}$ (otherwise, we would have detected it in Step 1). Suppose $j \geq 2$. Since iteration $j - 1$ is complete, by the previous claim $\mathcal{P}_c$ must preserve E_{j-1}. □

Armed with these claims, we can show that *Update*(i) can never fail and cannot run forever. Hence, it must output success.

Suppose it fails. Then it must have completed the iteration numbered lim (for whatever its current value). Note that by definition of lim, $E_{\leq lim} = [i] \cup \{k\}$. By Claim 6.10, all partitions in $\mathfrak{P}'$ must preserve $E_{\leq lim} = [i] \cup \{k\}$. So the only partitions that split $E_{\leq lim}$ are the labeled ones, which are at most $i-1$ is number. This contradicts the splitting property.

Suppose *Update*(i) does not terminate. Then it makes infinitely many promotions, and definitely at least $> (i+2)^2$ of them. At some stage, an element must be added to E_{i+1}, so the ith iteration of Step 2 is reached. (This means that $(i-1)$ iteration is completed.) Consider the situation just before this element is moved. During this iteration, some E_a must be empty for some $1 \leq a \leq i$. This is because $\{E_1, E_2, \ldots, E_i\}$ form a division of $[i-1]$. So C_a is also empty. Consider the set $E_{\leq a-1}$. By Claim 6.10, all partitions in $\mathfrak{P}'$ preserve $E_{\leq i-1} \supseteq E_{\leq a-1}$. By Claim 6.11, for $l \geq a+1$, all partitions in $\mathcal{C}_l$ preserve $E_{\leq a-1}$. The only partitions that split $E_{\leq a-1}$ must be $\mathcal{C}_{\leq a} = \mathcal{C}_{\leq a-1}$. But $|C_{\leq a-1}| = |E_{\leq a-1}| - 2$ (the difference of two occurs because we do not consider $\{i,k\}$ in $C_{\leq a-1}$), contradicting the splitting property. □

6.4. The general case

Now, we deal with simple, minimal identities and remove the strong minimality condition. This will come at a cost of an extra k factor in the rank bound. First, we recall the definition of gcd and simple parts of a general $\Sigma\Pi\Sigma$ circuit, as given in older works [Dvir and Shpilka 2006; Saxena and Seshadhri 2011a].

Definition 6.12 (*Gcd & Simple part*). Let $C = \sum_{i\in[k]} T_i$ be a $\Sigma\Pi\Sigma(k,d)$ circuit over a field $\mathbb{F}$. The *gcd* of C is defined to be the usual gcd of the polynomials T_i's, i.e. $\gcd(C) := \gcd(T_i | i \in [k])$.

The *simple* part of C is the $\Sigma\Pi\Sigma(k,d')$ circuit, $\mathrm{sim}(C) := C/\gcd(C)$, where $d' := d - \deg(\gcd(C))$.

Theorem 2.11 will be shown to be a consequence of Theorem 2.9. We prove that when $|\mathbb{F}| > d$, the rank of a simple, minimal $\Sigma\Pi\Sigma(k,d)$, independent-fanin k', identity is at most $2k^2 + (k-k') \cdot \mathrm{SG}_{k'}(\mathbb{F},d)$.

PROOF. (of Theorem 2.11) Let circuit C be $T_1 + \cdots + T_k = 0$. Wlog let $T_1, \cdots, T_{k'}$ be a linear basis for $T_1, \cdots, T_k$. Obviously, we have $1 < k' < k$ (first by simplicity and second by zeroness). By Theorem 2.4, there exists a nucleus K wrt the set $\mathcal{I} := [k']$. The rank of K is at most $2k^2$. So, it remains to bound the non-nucleus rank of C by $(k-k') \cdot \mathrm{SG}_{k'}(\mathbb{F},d)$.

As $T_1, \cdots, T_{k'}$ form a basis, for each $i \in [k'+1, k]$, there exists $\alpha_{i,j}$'s in $\mathbb{F}$ such that we have a zero circuit $D_i := \sum_{j\in[k']} \alpha_{i,j} T_j + T_i = 0$. Define N_i to be the set of j's for which $\alpha_{i,j} \neq 0$. Thus,

$$\forall i \in [k'+1, k],\ D_i = \sum_{j \in N_i} \alpha_{i,j} T_j + T_i = 0 \tag{4}$$

Since $\{\alpha_{i,j} T_j \mid j \in N_i\}$ are $|N_i|$ linearly independent terms, we get that D_i is a strongly minimal $\Sigma\Pi\Sigma(|N_i|+1, d)$ identity, for all $i \in [k'+1, k]$. By nucleus properties (recall Theorem 2.4-(3)), $\{K_j | j \in N_i\}$ are linearly independent polynomials (where $K_j = M(L_K(T_j))$), implying that the polynomials $\{K_j/g_i | j \in N_i\}$ are also linearly independent, where $g_i := M(L_K(\gcd(D_i)))$. Thus, the linear space K remains a nucleus of the new identity $\mathrm{sim}(D_i)$, showing at the same time that it is strongly minimal. We conclude that $\mathrm{sim}(D_i)$ is a simple, strongly minimal $\Sigma\Pi\Sigma(k_i, d_i)$ identity with nucleus

K (although of rk $< 2k^2$), $k_i \leq (k'+1), d_i \leq d$, for all $i \in [k'+1,k]$. Theorem 2.9 bounds the non-nucleus (non-K to be precise) rank of each of these identities by $\mathrm{SG}_{k'}(\mathbb{F}, d)$.

Define, $S := \bigcup_{i\in[k'+1,k]} N_i \subseteq [k']$. We first argue that $S = [k']$. Suppose not, so $S \subsetneq [k']$. By summing over i in Equation (4), $\sum_{i\in[k'+1,k]} T_i = \sum_{s\in S} \beta_s T_s$, for some β's in $\mathbb{F}$. Substituting this in the equation $C = 0$ we get,

$$C = C_{[k']} + C_{[k'+1,k]} \;=\; \sum_{i\in[k']} T_i + \sum_{s\in S} \beta_s T_s \;=\; 0.$$

As S is a proper subset of $[k']$, the above equation could only mean that a nontrivial combination of T_i ($i \in [k']$) is vanishing, contradicting the linear independence of those polynomials. Hence, $S = [k']$.

Now, suppose a linear form $\ell | \gcd(D_i)$ for all $i \in [k'+1,k]$. Then ℓ divides T_j for all $j \in \bigcup_{i\in[k'+1,k]} N_i \cup [k'+1,k] = [k]$ (since $S = [k']$). This means that ℓ divides every term in C, contradicting simplicity.

Thus, we conclude that each linear form ℓ of C does not divide some $\gcd(D_i)$. It appears in at least one of the circuits $\{\mathrm{sim}(D_i) | i \in [k'+1,k]\}$, whose total non-nucleus rank we have already bounded by $(k-k') \cdot \mathrm{SG}_{k'}(\mathbb{F}, d)$. That completes the proof. □

7. SYLVESTER-GALLAI RANK BOUNDS FOR ANY $\mathbb{F}$

We wish to bound $SG_k(\mathbb{F}, m)$, for *any* field $\mathbb{F}$. We will prove the following theorem, which can be seen as the first Sylvester-Gallai Theorem holding for *all* fields. A set of vectors in the projective space $\mathbb{FP}^n$ can be thought of as a *multiple-free* set of vectors S in $\mathbb{F}^{n+1}$. This means that no two vectors in S are scalar multiples of each other. The proof we present was given by [Saks 2010] (in our opinion, truly a proof from The Book [Erdös]). It is far more elegant and yields a much better constant factor that our original proof.

In some sense, bounds for $\mathrm{SG}_2(\mathbb{F}, m)$ are already implicit in known theorems (used to prove lower bounds for LDCs). Concretely, Corollary 2.9 of [Dvir and Shpilka 2006] can be interpreted as a proof that $\mathrm{SG}_2(\mathbb{F}, m) = O(\log m)$. This is an extension of theorems in [Goldreich et al. 2002] that prove this for $\mathbb{F}_2$. In the context of SG_2, these proofs can be interpreted as a "doubling trick". In essence, each time we want to increase the rank of an SG_2-closed set by 1, we are forced to double the number of vectors.

THEOREM 7.1 (HIGH DIMENSION SYLVESTER-GALLAI FOR ANY FIELD). *Suppose $k \in \mathbb{N}^{>1}$ and S is an SG_k-closed set of vectors in $\mathbb{FP}^n$ and rank $r \geq 1$. Then, $|S| \geq 2^{(r/(k-1))-1}$. In other words, for every $m \in \mathbb{N}^{>1}$, $SG_k(\mathbb{F}, m) \leq (k-1)\lg 2m$.*

PROOF. Let $F(r,k) := 2^{(r/(k-1))-1}$, where $k \geq 2$ and $r \geq 1$. We prove by induction on lexicographic order on (r,k) that a multiple-free SG_k-closed set of rank r has size at least $F(r,k)$. For the base case, note that when $r \leq k-1$, then $F(r,k) \leq 1$. Since $|S| \geq r \geq 1$, the bound is true. So, we assume that $r \geq k$.

If S is SG_{k-1}-closed, then we are done by the induction hypothesis. This is because $|S| \geq F(r,k-1)$ and $F(r,k-1) \geq F(r,k)$. So, we can assume that S is not SG_{k-1}-closed, and there exists a linearly independent set of vectors $v_1, v_2, \ldots, v_{k-1} \in S$ that span no other vector in S. (Note that we include the case $k = 2$ in this proof.) Since the rank is r, we extend this to a set of r linearly independent vectors in S, denoted by $v_1, v_2, \ldots, v_r$. Let T be the subset of S spanned by $v_k, \ldots, v_r$. Since S is SG_k-closed, T is also SG_k-closed. The rank of T is $r-(k-1) < r$, so we can apply the induction hypothesis. This yields that $|T| \geq F(r-(k-1),k)$.

For each $v \in T$, the set $v, v_1, v_2, \ldots, v_{k-1}$ must span another vector in S. Call this vector $f(v)$, which must be a non-trivial combination of the above vectors. The vector $f(v)$ cannot lie in T, because this would imply that some linear combination of

$v_1, v_2, \ldots, v_{k-1}$ lies in T. But T was chosen to be independent to all these vectors. Similarly, vector $f(v)$ cannot lie in the span of $v_1, v_2, \ldots, v_{k-1}$. Hence, $f(v)$ can be expressed as $\alpha_v v + \sum_{i \le k-1} \beta_{v,i} v_i$, where $\alpha_v \neq 0$ and some $\beta_{v,i} \neq 0$.

Suppose for $v \neq v'$ $(v, v' \in T)$, $f(v) = f(v')$. Then $\alpha_v v - \alpha_{v'} v' = \sum_{i \le k-1} (\beta_{v',i} - \beta_{v,i}) v_i$. Since both $\alpha_v, \alpha_{v'} \neq 0$ and S is multiple-free, the left hand side is a non-zero vector spanned by T. But the right hand side is independent to T or is zero. Therefore, by contradiction, we deduce that for $v \neq v'$, $f(v) \neq f(v')$.

We can bound $|S| \geq 2|T| \geq 2F(r-(k-1), k)$. This is $2^{(r-k+1)/(k-1)} = 2^{(r/(k-1))-1} = F(r,k)$. □

We give a simple construction providing a lower bound for $\mathrm{SG}_k(\mathbb{F}_p, m)$.

THEOREM 7.2. *For every $k \geq 2, d \geq 2$, and prime p, there exists a (multiple-free) set of $(k-1)(p^d-1)/(p-1)$ vectors of rank $(k-1)d$ that form an SG_k-closed set. In other words, $\mathrm{SG}_k(\mathbb{F}_p, m) \geq (k-1)\log_p(m/(k-1))$.*

PROOF. We will construct vectors over $\mathbb{F}_p^n$, where $n = (k-1)d$. Think of the coordinates as broken into $(k-1)$ contiguous blocks, where each block has d coordinates. The final set S will comprise of $(k-1)$ subsets $S_1, S_2, \ldots, S_{k-1}$. The vectors in subset S_i will have non-zero coordinates only in the ith block of coordinates. So we can think of each S_i as vectors in $\mathbb{F}_p^d$, with each S_i being defined over disjoint sets of coordinates. The set S_i will just be the multiple-free, maximal subset of $\mathbb{F}_p^d \setminus \{0\}$. Naturally, each S_i is SG_2-closed.

The overall set S has rank $r := (k-1)d$ and size $m := (k-1)(p^d-1)/(p-1)$. Consider any k vectors in S. By the pigeonhole-principle, two of these vectors must lie in the same set (say) S_i. Since S_i is SG_2-closed, there is a non-trivial combination of these vectors inside S_i. Hence, S is SG_k-closed. □

8. PROOFS FOR ALGEBRAIC LEMMAS

For convenience, we restate the lemmas before the proofs.

Lemma 3.2 (restated). *Let $f_1, \ldots, f_m$ be multiplication terms generating an ideal I, let $\ell \in L(\mathcal{R})$ and $g \in \mathcal{R}$. If $\ell \notin \mathrm{radsp}(I)$ then: $\ell g \in I$ iff $g \in I$.*

PROOF. Assume $\ell \notin \mathrm{radsp}(I)$. If $I = \{0\}$ then the lemma is of course true. So let us assume that $I \neq \{0\}$ and $\mathrm{rk}(\mathrm{radsp}(I)) =: r \in [n-1]$. As $\ell \notin \mathrm{radsp}(I)$ there exists an invertible linear transformation $\tau : L(\mathcal{R}) \to L(\mathcal{R})$ that maps each form of $\mathrm{radsp}(I)$ to $\mathrm{sp}(x_1, \ldots, x_r)$ and maps ℓ to x_n. Now suppose that $\ell g \in I$. This means that there are $q_1, \ldots, q_m \in R$ such that $\ell g = \sum_{i=1}^m q_i f_i$. Apply τ on this to get:

$$x_n g' = \sum_{i=1}^m q_i' \tau(f_i). \tag{5}$$

We know that $\tau(f_i)$'s are free of x_n. Express g', q_i'-s as polynomials wrt x_n, say

$$g' = \sum_{j \geq 0} a_j x_n^j, \text{ where } a_j \in \mathbb{F}[x_1, \ldots, x_{n-1}] \tag{6}$$

$$q_i' = \sum_{j \geq 0} b_{i,j} x_n^j, \text{ where } b_{i,j} \in \mathbb{F}[x_1, \ldots, x_{n-1}] \tag{7}$$

Now for some $d \geq 1$ compare the coefficients of x_n^d on both sides of Equation (5). We get $a_{d-1} = \sum_{i=1}^m b_{i,d} \tau(f_i)$, thus a_{d-1} and $a_{d-1} x_n^{d-1}$ are in $\langle \tau(f_1), \ldots, \tau(f_m) \rangle$. Doing this

for all $d \geq 1$, we get $g' \in \langle \tau(f_1), \ldots, \tau(f_m) \rangle$, hence $g = \tau^{-1}(g') \in \langle f_1, \ldots, f_m \rangle = I$. This finishes the proof. □

Lemma 3.3 (restated). *Suppose $f_1, \ldots, f_m, g$ are homogeneous polynomials in $\mathcal{R}$. Then,*
1) If deg$(g) <$ deg(f_m) then: $g \in \langle f_1, \ldots, f_m \rangle$ iff $g \in \langle f_1, \ldots, f_{m-1} \rangle$.
2) If deg$(g) =$ deg(f_m) then: $g \in \langle f_1, \ldots, f_m \rangle$ iff $\exists a \in \mathbb{F}$, $(g + af_m) \in \langle f_1, \ldots, f_{m-1} \rangle$.

PROOF. Say, $g \in \langle f_1, \ldots, f_m \rangle$. Then, by definition, there exist q's in $\mathcal{R}$ such that,

$$g = \sum_{i=1}^{m} q_i f_i. \tag{8}$$

Let $d :=$ deg(g). If we compare the monomials of degree d on both sides of Equation (8) then the LHS gives g. In the RHS we see that an f_i of degree d_i contributes $[q_i]_{(d-d_i)} f_i$, where $[q]_j$ is defined to be the sum of the degree j terms of q (and, zero if $j < 0$). Thus, $g = \sum_{i=1}^{m} [q_i]_{d-d_i} f_i$. This equation proves both the properties at once. □

Lemma 3.5 (restated). *Let I be an ideal generated by multiplication terms $\{f_1, \ldots, f_m\}$ and define $U :=$ radsp(I). Let g, h be multiplication terms such that $g \equiv h \not\equiv 0$ (mod I). Then there is a U-matching between $L_U(g), L_U(h)$ and one between $L_U^c(g), L_U^c(h)$.*

PROOF. Define $g_0 := M(L_U(g))$ and $h_0 := M(L_U(h))$. Suppose the list $L_U(g)$ is larger than the list $L_U(h)$. By the congruence we have $h \in \langle I, g_0 \rangle$. Observe that radsp$(I, g_0) = U$. Suppose some form $\ell \notin U$ divides h, so $h = \ell h'$. By Lemma 3.2, $h' \in \langle I, g_0 \rangle$. Applying inductively for all non-U forms of h, we deduce that $h_0 \in \langle I, g_0 \rangle$. As $\langle I, g_0 \rangle$ is a homogeneous ideal and deg$(h_0) <$ deg(g_0) we get by Lemma 3.3 that $h_0 \in I$. But this means $h \in I$, which contradicts the hypothesis. Thus, deg$(h_0) \geq$ deg(g_0) and by symmetry we get them infact equal. Thus, the lists $L_U(g), L_U(h)$ are of equal size, which trivially U-matches them.

We will show that for any $\ell \in L(\mathcal{R}) \setminus U$, the number of forms that are similar to ℓ mod U in $L_U^c(g)$ is equal to that in $L_U^c(h)$. This fact will prove the lemma as it shows that every form in $L_U^c(g)$ can be U-matched to a distinct form in $L_U^c(h)$.

Pick an $\ell \in L(\mathcal{R}) \setminus U$. Let g_1 be the product of the forms that are similar to ℓ mod U in $L_U^c(g)$ (if none exist then set $g_1 = 1$), similarly define h_1 from h. Suppose deg$(h_1) <$ deg$(g_1) =: d$. By the congruence we have $h \in \langle I, g_0 g_1 \rangle$. Observe that radsp$(I, g_0 g_1) = U \oplus \mathbb{F}\ell$. By the inductive argument given above using Lemma 3.2, we can drop the non sp(U, ℓ) forms of h to get

$$h_0 h_1 \in \langle I, g_0 g_1 \rangle. \tag{9}$$

As we have already shown deg$(h_0) =$ deg(g_0), we have deg$(h_0 h_1) <$ deg$(g_0 g_1)$. Thus, by Lemma 3.3, the above Equation entails $h_0 h_1 \in I$. So $h \in I$, contradicting the hypothesis. This shows the number of forms that are similar to ℓ mod U in $L_U^c(g)$ is equal to that in $L_U^c(h)$, finishing the proof. □

To prove Lemma 5.2, we need a metric associated with matchings, first introduced in Section 2.4.2 of [Saxena and Seshadhri 2011a]. It is essentially the factor by which a matching scales-up a linear form (modulo the ideal).

Definition 8.1 (*Scaling factor*). Let K be a subspace of $L(\mathcal{R})$ and L_1, L_2 be two lists of linear forms in $L(\mathcal{R}) \setminus K$. Let π be a K-matching between L_1, L_2. Then for every $\ell \in L_1$, there is a *unique* $c_\ell \in \mathbb{F}^*$ such that $\pi(\ell) \in c_\ell \ell + K$ (if there is another $d \in \mathbb{F}$ with $\pi(\ell) \in d\ell + K$, then $(c_\ell - d)\ell \in K$, implying $\ell \in K$, a contradiction).

We define the *scaling factor of* π, $\mathrm{sc}(\pi) := \prod_{\ell \in L_1} c_\ell$.

Lemma 5.2 (restated). *Suppose $C = \sum_i T_i$ is a $\Sigma\Pi\Sigma(k,d)$ identity and K is a subspace of $L(\mathcal{R})$ such that T_1, T_i are K-matched, for all $i \in [k]$. Then the terms $M(L_K(T_i))$, for $i \in [k]$, are all of the same degree, say d', and form a $\Sigma\Pi\Sigma(k,d')$ identity $\sum_{i\in[k]} \alpha_i M(L_K(T_i))$, for some $\alpha_i \in \mathbb{F}^*$.*

PROOF. Since T_1, T_i are K-matched, we get from the definition of matching that terms $M(L_K(T_1))$, $M(L_K(T_i))$ have the same degree $d' \geq 0$. Furthermore, $M(L^c_K(T_1))$ and $M(L^c_K(T_i))$ are also K-matched, call this induced matching π_i. As all the forms in $L^c_K(T_1)$ are outside K, the scaling factor $\mathrm{sc}(\pi_i)$ is well defined, for all $i \in [k]$.

Fix a subspace U such that $L(\mathcal{R}) = K \oplus U$ and let $r := \mathrm{rk}(K)$. Fix an invertible linear transformation $\tau : L(\mathcal{R}) \to L(\mathcal{R})$ that maps K to $\mathrm{sp}(x_1, \ldots, x_r)$. It follows that for any form $\ell \in L^c_K(T_1)$, $\tau(\ell)$ is a form with a nonzero coefficient wrt some x_i, $i > r$ (otherwise $\tau(\ell) \in \mathrm{sp}(x_1, \ldots, x_r)$, thus $\ell \in K$, a contradiction). Call the largest such i, j_ℓ. If we look at the product (note: it is over a list so it could have repeated factors),

$$\alpha_1 := \prod_{\ell \in L^c_K(T_1)} [x_{j_\ell}]\tau(\ell) \tag{10}$$

($[\overline{x}^{\bar{i}}]f$ gives the coefficient of the monomial $\overline{x}^{\bar{i}}$ in f), then it is the coefficient of $\prod_{\ell \in L^c_K(T_1)} x_{j_\ell}$ in $\tau(M(L^c_K(T_1)))$, in other words, α_1 is its leading coefficient wrt lexicographic ordering of variables. Note that, for $i \in [k]$, π_i still $\tau(K)$-matches $\tau(L^c_K(T_1)), \tau(L^c_K(T_i))$ with the same scaling factor (if $\pi_i(\ell) \in c_\ell \ell + K$ then $\tau(\pi_i(\ell)) \in c_\ell \tau(\ell) + \tau(K)$). This means that the leading coefficient of $\tau(M(L^c_K(T_i)))$ is $\mathrm{sc}(\pi_i) \cdot \alpha_1 =: \alpha_i$, for all $i > 1$. Thus, we have pinpointed the coefficient of $\prod_{\ell \in L^c_K(T_1)} x_{j_\ell}$ in $\tau(M(L^c_K(T_i)))$ as α_i, for all $i \in [k]$. Now compare the coefficients of $\prod_{\ell \in L^c_K(T_1)} x_{j_\ell}$ in the identity $\tau(C) = 0$. This gives $\sum_{i\in[k]} \alpha_i \cdot \tau(M(L_K(T_i))) = 0$. Applying the inverse of τ, we get the nucleus identity. □

Lemma 6.3 (restated). *Let $|\mathbb{F}| > d$ and C be a $\Sigma\Pi\Sigma(k,d)$ identity, over $\mathbb{F}$, with nucleus K. Let $y_0 \in L(\mathcal{R})^*$ and U be a subspace of $L(\mathcal{R})$ such that $L(\mathcal{R}) = \mathbb{F}y_0 \oplus U \oplus K$. Then there exists an invertible linear transformation $\tau : L(\mathcal{R}) \to L(\mathcal{R})$ that fixes K and :*

1) *$\tau(C)$ is also a $\Sigma\Pi\Sigma(k,d)$ identity with nucleus K and the same simplicity, minimality properties.*
2) *Every form in $L^c_K(\tau(C)) = \tau(L^c_K(C))$ is monic wrt y_0.*

PROOF. Let $r := \mathrm{rk}(\mathbb{F}y_0 \oplus U)$. Fix a basis $\{y_0, \ldots, y_{r-1}\}$ of $\mathbb{F}y_0 \oplus U$. (Think of each y_i as a column vector.) Let $\overline{y}$ denote the matrix $[y_0, \cdots, y_{r-1}]^T$, where each row is a linear form. Let $\ell \in L^c_K(T_1)$. Then there is a unique nonzero (column) vector $\overline{\alpha_\ell} \in \mathbb{F}^r$ and a $v_\ell \in K$, such that $\ell = \overline{\alpha_\ell}^T \cdot \overline{y} + v_\ell$. We intend τ to be a linear transformation that fixes each element in K and maps $\overline{y}$ to $A\overline{y}$ where $A \in \mathbb{F}^{r\times r}$. Such a τ will map ℓ to $\tau(\overline{\alpha_\ell}^T \cdot \overline{y}) + v_\ell = \overline{\alpha_\ell}^T \cdot \tau(\overline{y}) + v_\ell = \overline{\alpha_\ell}^T A\overline{y} + v_\ell$. To make $\tau(\ell)$ monic in y_0 we need to choose A such that the first coordinate in $\overline{\alpha_\ell}^T A$ is nonzero, i.e. $\overline{\alpha_\ell}^T A_{*1} \neq 0$ where A_{*1} is the first column of A. Thus, we want an A such that $\prod_{\ell \in L^c_K(T_1)} \overline{\alpha_\ell}^T A_{*1} \neq 0$.

Now the nonzero multivariate polynomial $f(\overline{Y}) := \prod_{\ell \in L^c_K(T_1)} \overline{\alpha_\ell}^T \overline{Y}$ has degree at most $d < |\mathbb{F}|$. Hence, by [Schwartz 1980; Zippel 1979; DeMillo and Lipton 1978] lemma there exists a point $\overline{Y} \in \mathbb{F}^r$ at which f is nonzero. We can fix A_{*1} to be that point. This fixes just one column of A to a nonzero vector and we can arbitrarily fix the rest such that A is an invertible matrix. Thus, the corresponding invertible τ makes each

$\ell \in L_K^c(T_1)$ monic in y_0. Since τ fixes the nucleus K, matching property of the nucleus tells us that every form in $L_K^c(\tau(C)) = \tau(L_K^c(C))$ is monic in y_0.

Since τ is an invertible linear transformation, it is actually an automorphism of $L(\mathcal{R})$ and, in particular, the zeroness, simplicity and minimality properties of C are invariant under it. □

In Lemma 3.2, we have seen a cancellation rule for non-zerodivisors. The following is a stronger form of cancellation.

Lemma 6.4 (restated). *Let K be some subspace of $L(\mathcal{R})$ and let $\ell_1, \ldots, \ell_m \in L(\mathcal{R}) \setminus K$ be linearly independent modulo K. Let $f_1, \ldots, f_m$ be multiplication terms similar to powers of $\ell_1, \ldots, \ell_m$ respectively modulo K (i.e. each form in f_i is in $(\mathbb{F}^*\ell_i + K)$). Let $\ell \in L(\mathcal{R})^*$ such that for some $s \in [m]$, $\ell \in \mathbb{F}\ell_s + K$. Then, for any polynomial $f \in \mathcal{R}$,*

$$\ell f \in \langle f_1, \ldots, f_m \rangle \text{ iff } f \in \langle f_1, \ldots, \frac{f_s}{\gcd(f_s, \ell)}, \ldots, f_m \rangle.$$

PROOF. Suppose $\ell f \in \langle f_1, \ldots, f_m \rangle$. Then, by definition, there exist q's in $\mathcal{R}$ such that,

$$\ell f = \sum_{i=1}^{m} q_i f_i. \tag{11}$$

Additionally assume these q_i's to be such that the set $J := \{j \in [m] \setminus \{s\} \mid \ell \nmid q_j\}$ is the smallest possible. If $\ell | q_i$, for all $i \in [m] \setminus \{s\}$, then ℓ has to divide $q_s f_s$. This means that ℓ has to divide $q_s \gcd(\ell, f_s)$, thus we get,

$$f = \sum_{i \in [m] \setminus \{s\}} \frac{q_i}{\ell} f_i \; + \; \frac{q_s \gcd(\ell, f_s)}{\ell} \cdot \frac{f_s}{\gcd(f_s, \ell)}$$

and we are done.

So the remaining case is when the set $J := \{j \in [m] \setminus \{s\} \mid \ell \nmid q_j\}$ is nonempty. Fix an element $j^* \in J$. Consider the ideal $I := \langle \{\ell, f_s\} \cup \{f_j | j^* \neq j \in J\} \rangle$. Reducing Equation (11) modulo I we get, $q_{j^*} f_{j^*} \equiv 0 (\text{mod } I)$. Note that $\text{radsp}(I) \subseteq K + \text{sp}(\{\ell_j | j^* \neq j \in [m]\})$ while each form in $L(f_{j^*})$ is in $(\mathbb{F}^*\ell_{j^*} + K)$ which is disjoint from $\text{radsp}(I)$. Thus by Lemma 3.2 we can drop f_{j^*} from the last congruence and get $q_{j^*} \in I$. This means $q_{j^*} f_{j^*} \in \langle \{\ell f_{j^*}, f_s\} \cup \{f_j | j^* \neq j \in J\} \rangle$. We plug this in the j^*-th summand of Equation (11) and after simplifications get (verify that the $[m] \setminus (\{s\} \cup J)$ summands are unaffected):

$$\begin{aligned} \ell f &= \sum_{i=1}^{m} q_i f_i \\ &= q'_s f_s + (\ell q'_{j^*}) f_{j^*} + \sum_{j \in J \setminus \{j^*\}} q'_j f_j + \sum_{j \in [m] \setminus (\{s\} \cup J)} q_j f_j \end{aligned}$$

Notice that for $j \in [m] \setminus (\{s\} \cup J)$, ℓ divides q_j, thus the above equation contradicts the assumed minimality of J. This shows that J was empty to begin with, thus finishing the proof. □

9. CONCLUSION

In this work we developed new methods to study depth-3 identities. These ideal methods hinge on a classification of zerodivisors of the ideals generated by gates of a $\Sigma\Pi\Sigma$ circuit (eg. Lemmas 3.2, 3.5 and 6.4). That is useful in proving an ideal version of

Chinese remaindering tailor-made for $\Sigma\Pi\Sigma$ circuits. As a byproduct, it shows the existence of a low rank *nucleus identity* C' inside *any* given $\Sigma\Pi\Sigma(k,d)$ identity C (when C is not minimal, C' can still be defined but it might not be homogeneous). The properties of the nucleus identity are an important part of an identity and it might be useful for PIT to understand (or classify) it further. Can the rank bound for the nucleus identity be improved to $O(k)$? More importantly, can the rank bound for simple minimal real $\Sigma\Pi\Sigma(k,d)$ identities be improved to $O(k)$? The best constructions known, since [Dvir and Shpilka 2006], have rank $4(k-2)$. Over other fields, our upper bound of $O(k^2 \log d)$ still leaves some gap in understanding the exact dependence on k. Of course, the most important question is whether our techniques can help construct a truly polynomial time deterministic (even non-blackbox) algorithm for PIT.

We generalize the notion of Sylvester-Gallai configurations to any field and define a parameter $\mathrm{SG}_k(\mathbb{F},m)$ associated with field $\mathbb{F}$. This number seems to be a fundamental property of a field, and as we show, is very closely related to $\Sigma\Pi\Sigma$ identities. It would be interesting to obtain bounds for $\mathrm{SG}_k(\mathbb{F},m)$ for different $\mathbb{F}$. For example, as also asked by [Kayal and Saraf 2009], can we nontrivially bound the number $\mathrm{SG}_k(\mathbb{F},m)$ for other interesting fields: $\mathbb{C}$, finite fields with large characteristic, or even p-adic fields? The only known SG_k rank bounds are those for $\mathbb{R}$, $\mathrm{SG}_2(\mathbb{C},m) \leq 3$, and $\mathrm{SG}_2(\mathbb{F},m) = O(\log m)$. We shed (a little) light on SG rank bounds by showing $\mathrm{SG}_k(\mathbb{F},m) = O(k \log m)$, which was subsequently improved for finite fields in [Bhattacharyya et al. 2011]. It would be interesting to generalize their bound of $\mathrm{SG}_2(\mathbb{F}_p,m) = O(\mathrm{poly}(p) + \log_p m)$ to one for $\mathrm{SG}_k(\mathbb{F}_p,m)$.

ACKNOWLEDGMENTS

We are grateful to Hausdorff Center for Mathematics, Bonn for the kind support, especially, hosting CS when part of the work was done. CS would also like to acknowledge the Early Career-LDRD program at Sandia National Labs for funding during writing the final version. We are grateful to Michael Saks for sharing his more elegant proof of Theorem 7.1. NS thanks Nils Frohberg for several detailed discussions that clarified the topic of incidence geometry and Sylvester-Gallai theorems. CS is extremely grateful to Ken Clarkson and especially to T. S. Jayram for their suggestions in improving the presentation. We also thank Malte Beecken, Johannes Mittmann and Thomas Thierauf for several interesting discussions.

REFERENCES

AGRAWAL, M. 2005. Proving lower bounds via pseudo-random generators. In *Proceedings of the 25th Annual Foundations of Software Technology and Theoretical Computer Science (FSTTCS)*. 92–105.

AGRAWAL, M. 2006. Determinant versus permanent. In *Proceedings of the 25th International Congress of Mathematicians (ICM)*. Vol. 3. 985–997.

AGRAWAL, M. AND BISWAS, S. 2003. Primality and identity testing via Chinese remaindering. *Journal of the ACM 50,* 4, 429–443. (Conference version in FOCS 1999).

AGRAWAL, M., SAHA, C., SAPTHARISHI, R., AND SAXENA, N. 2012. Jacobian hits circuits: Hitting- sets, lower bounds for depth-D occur-k formulas & depth-3 transcendence degree-k circuits. In *STOC*. 599–614.

AGRAWAL, M. AND SAPTHARISHI, R. 2009. Classifying polynomials and identity testing. In *Current Trends in Science, Indian Academy of Sciences*. 149–162.

AGRAWAL, M. AND VINAY, V. 2008. Arithmetic circuits: A chasm at depth four. In *Proceedings of the 49th Annual Symposium on Foundations of Computer Science (FOCS)*. 67–75.

ANDERSON, M., VAN MELKEBEEK, D., AND VOLKOVICH, I. 2011. Derandomizing polynomial identity testing for multilinear constant-read formulae. In *Proceedings of the 26th Conference on Computational Complexity (CCC)*. 273–282.

ARVIND, V. AND MUKHOPADHYAY, P. 2010. The ideal membership problem and polynomial identity testing. *Inf. Comput. 208,* 4, 351–363. (Conference version in ISAAC 2007).

BEECKEN, M., MITTMANN, J., AND SAXENA, N. 2011. Algebraic independence and blackbox identity testing. In *Proceedings of the 38th Annual International Colloquium on Automata, Languages and Programming (ICALP)*. 137–148.

BHATTACHARYYA, A., DVIR, Z., SARAF, S., AND SHPILKA, A. 2011. Tight lower bounds for 2-query LCCs over finite fields. In *Proceedings of the 52nd Annual Symposium on Foundations of Computer Science (FOCS)*. 638–647.

BONNICE, W. AND EDELSTEIN, M. 1967. Flats associated with finite sets in P^d. *Niew. Arch. Wisk. 15*, 11–14.

BORWEIN, P. AND MOSER, W. O. J. 1990. A survey of Sylvesters problem and its generalizations. *Aequationes Mathematicae 40,* 1, 111–135.

CHEN, Z. AND KAO, M. 2000. Reducing randomness via irrational numbers. *SIAM J. on Computing 29,* 4, 1247–1256. (Conference version in STOC 1997).

DEMILLO, R. A. AND LIPTON, R. J. 1978. A Probabilistic Remark on Algebraic Program Testing. *Inf. Process. Lett. 7,* 4, 193–195.

DVIR, Z. AND SHPILKA, A. 2006. Locally decodable codes with 2 queries and polynomial identity testing for depth 3 circuits. *SIAM J. on Computing 36,* 5, 1404–1434. (Conference version in STOC 2005).

ELKIES, N. D., PRETORIUS, L. M., AND SWANEPOEL, C. J. 2006. Sylvester-Gallai theorems for complex numbers and quaternions. *Discrete & Computational Geometry 35,* 3, 361–373.

ERDÖS, P. `http://en.wikipedia.org/wiki/Paul_Erd%C5%91s`.

GOLDREICH, O., KARLOFF, H., SCHULMAN, L., AND TREVISAN, L. 2002. Lower bounds for linear locally decodable codes and private information retrieval. In *Proceedings of the 17th Annual Computational Complexity Conference (CCC)*. 175–183.

HANSEN, S. 1965. A generalization of a theorem of Sylvester on the lines determined by a finite point set. *Mathematica Scandinavia 16*, 175–180.

HEINTZ, J. AND SCHNORR, C. P. 1980. Testing polynomials which are easy to compute (extended abstract). In *Proceedings of the twelfth annual ACM Symposium on Theory of Computing*. New York, NY, USA, 262–272.

KABANETS, V. AND IMPAGLIAZZO, R. 2004. Derandomizing polynomial identity tests means proving circuit lower bounds. *Computational Complexity 13,* 1, 1–46. (Conference version in STOC 2003).

KARNIN, Z., MUKHOPADHYAY, P., SHPILKA, A., AND VOLKOVICH, I. 2010. Deterministic identity testing of depth-4 multilinear circuits with bounded top fan-in. In *Proceedings of the 42nd ACM Symposium on Theory of Computing (STOC)*. 649–658.

KARNIN, Z. AND SHPILKA, A. 2008. Deterministic black box polynomial identity testing of depth-3 arithmetic circuits with bounded top fan-in. In *Proceedings of the 23rd Annual Conference on Computational Complexity (CCC)*. 280–291.

KARNIN, Z. S. AND SHPILKA, A. 2009. Reconstruction of generalized depth-3 arithmetic circuits with bounded top fan-in. In *Proceedings of the 24th Annual Conference on Computational Complexity (CCC)*. 274–285.

KAYAL, N. AND SARAF, S. 2009. Blackbox polynomial identity testing for depth 3 circuits. In *Proceedings of the 50th Annual Symposium on Foundations of Computer Science (FOCS)*. 198–207.

KAYAL, N. AND SAXENA, N. 2007. Polynomial identity testing for depth 3 circuits. *Computational Complexity 16,* 2, 115–138. (Conference version in CCC 2006).

KLIVANS, A. AND SPIELMAN, D. A. 2001. Randomness efficient identity testing of multivariate polynomials. In *Proceedings of the 33rd Symposium on Theory of Computing (STOC)*. 216–223.

LEWIN, D. AND VADHAN, S. 1998. Checking polynomial identities over any field: Towards a derandomization? In *Proceedings of the 30th Annual Symposium on the Theory of Computing (STOC)*. 428–437.

MULMULEY, K. 2012. Geometric Complexity Theory V: Equivalence between blackbox derandomization of polynomial identity testing and derandomization of Noether's Normalization Lemma. In *FOCS*.

MULMULEY, K. D. 2011. On P vs. NP and geometric complexity theory: Dedicated to Sri Ramakrishna. *J. ACM 58,* 2, 5:1–5:26.

RAZ, R. 2010. Tensor-rank and lower bounds for arithmetic formulas. In *Proceedings of the 42nd Annual ACM Symposium on Theory of Computing (STOC)*. 659–666.

SAHA, C., SAPTHARISHI, R., AND SAXENA, N. 2012. A case of depth-3 identity testing, sparse factorization and duality. *Comp. Complex.*. ECCC TR11-021.

SAKS, M. 2010. Personal communication.

SARAF, S. AND VOLKOVICH, I. 2011. Black-box identity testing of depth-4 multilinear circuits. In *Proceedings of the 43rd ACM Symposium on Theory of Computing (STOC)*. 421–430.

SAXENA, N. 2008. Diagonal circuit identity testing and lower bounds. In *Proceedings of the 35th Annual International Colloquium on Automata, Languages and Programming (ICALP)*. 60–71.

SAXENA, N. 2009. Progress on polynomial identity testing. *Bulletin of the European Association for Theoretical Computer Science (EATCS)- Computational Complexity Column* 99, 49–79.

SAXENA, N. AND SESHADHRI, C. 2010. From Sylvester-Gallai configurations to rank bounds: Improved black-box identity test for depth-3 circuits. In *Proceedings of the 51st Annual IEEE Symposium on Foundations of Computer Science (FOCS)*. 21–29.

SAXENA, N. AND SESHADHRI, C. 2011a. An Almost Optimal Rank Bound for Depth-3 Identities. *SIAM J. Comp. 40,* 1, 200–224. (Conference version in CCC 2009).

SAXENA, N. AND SESHADHRI, C. 2011b. Blackbox identity testing for bounded top fanin depth-3 circuits: the field doesn't matter. In *Proceedings of the 43rd annual ACM Symposium on Theory of Computing (STOC)*. 431–440.

SCHWARTZ, J. T. 1980. Fast probabilistic algorithms for verification of polynomial identities. *Journal of the ACM 27,* 4, 701–717.

SHPILKA, A. 2009. Interpolation of depth-3 arithmetic circuits with two multiplication gates. *SIAM J. Comput. 38,* 6, 2130–2161. (Conference version in STOC 2007).

SHPILKA, A. AND VOLKOVICH, I. 2008. Read-once polynomial identity testing. In *Proceedings of the 40th Annual ACM Symposium on Theory of Computing (STOC)*. 507–516.

SHPILKA, A. AND VOLKOVICH, I. 2009. Improved polynomial identity testing for read-once formulas. In *Proceedings of the 13th International Workshop on Randomization and Computation (RANDOM)*. 700–713.

SHPILKA, A. AND YEHUDAYOFF, A. 2010. Arithmetic Circuits: A survey of recent results and open questions. *Foundations and Trends in Theoretical Computer Science 5,* 3-4, 207–388.

ZIPPEL, R. 1979. Probabilistic algorithms for sparse polynomials. In *Proceedings of the International Symposium on Symbolic and Algebraic Manipulation (EUROSAM)*. 216–226.

Received January 2012; revised October 2012; accepted -